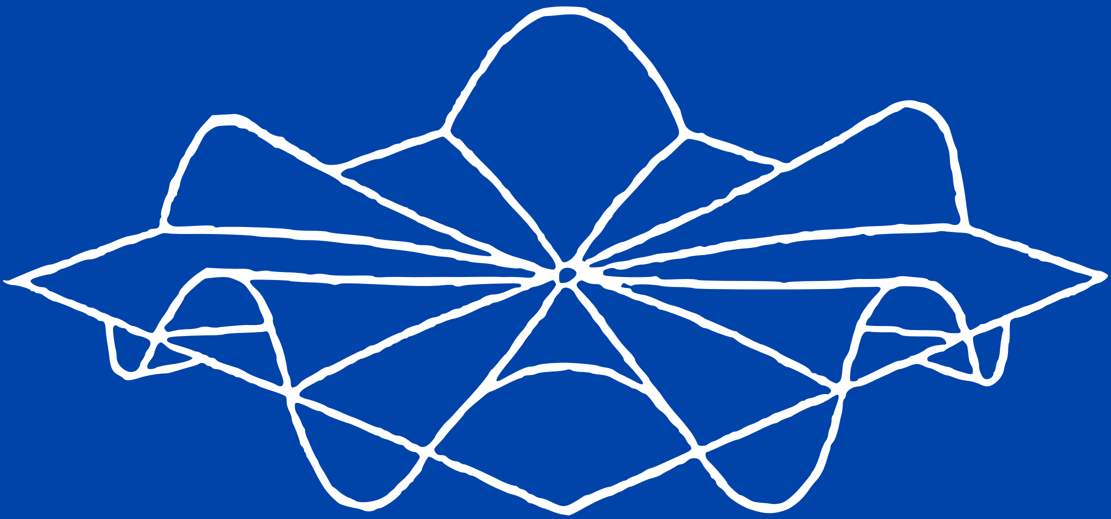
V.A. Ilyin, E.G. Poznyak

FUNDAMENTALS OF MATHEMATICAL ANALYSIS



Mir Publishers • Moscow

В. А. ИЛЬИН, Э. Г. ПОЗНЯК

ОСНОВЫ МАТЕМАТИЧЕСКОГО АНАЛИЗА

Часть II

Издательство «Наука» Москва

V. A. ILYIN and E. G. POZNYAK

Fundamentals of Mathematical Analysis

First published 1982
Revised from 1980 Russian edition

На английском языке

PREFACE

This book is based on the lectures read by authors at Moscow State University for a number of years.

As in Part 1, the authors strived to make presentation systematic and to set off the most important notions and theorems.

Besides the basic curriculum material, this book contains some additional questions that play an important part in various branches of modern mathematics and physics (the theory of measure and Lebesgue integrals, the theory of Hilbert spaces and of self-adjoint linear operators in these spaces, questions of regularization of Fourier series, the theory of differential forms in Euclidean spaces, etc.). Some of the topics, such as the conditions for termwise differentiation and termwise integration of functional sequences and functional series, the theorem on the change of variables in a multiple integral, Green's and Stokes's formulas, necessary conditions for a bounded function to be integrable in the sense of Riemann and in the sense of Lebesgue, are treated more generally and under weaker assumptions than usual.

As in Part 1, we discuss in this book some questions related to computational mathematics, including first of all approximate calculation of multiple integrals in the supplement to Chapter 2 and calculation of the values of functions from the approximate values of Fourier coefficients (A.N. Tichonoff's regularization method) in the Appendix.

The material of this book, together with that of Part 1 published earlier, constitutes an entire university course in mathematical analysis.

Note that throughout this text Part 1 is referred to as Volume 1 and designated [1]. It should also be stressed that when reading this book Chapter 8, The Lebesgue Integral and Measure, Chapter 11, Hilbert Space, and all the supplements may be skipped without impairing the understanding of the rest of the text.

The authors feel deeply indebted to A.N. Tichonoff and A.G. Sveshnikov for much valuable advice and numerous profound criticisms, to Sh.A. Alimov, who has done more than just editing this book, to L.D. Kudryavtsev and S.A. Lomov for a great number of valuable criticisms, to P.S. Modenov and Ya.M. Zhileikin, who have made available to the authors materials on field theory and approximate methods of evaluating multiple integrals.

V. Ilyin, E. Poznyak

CONTENTS

CHAPTER 1

FUNCTIONAL SEQUENCES AND FUNCTIONAL SERIES

In this chapter we shall study sequences and series whose members are not numbers but functions defined on some given set. Such sequences and series are widely used to represent the functions and to compute them approximately.

1.1. UNIFORM CONVERGENCE

1.1.1. The functional sequence and the functional series. *Let* $\{x\}$ *be some set*. Then, if we assign to each* n *of the natural numbers* $1, 2, \ldots$ $\ldots, n, \ldots$ *by a definite rule some function* $f_n(x)$ *defined on* $\{x\}$, *the set of the numbered functions* $f_1(x), f_2(x), \ldots, f_n(x), \ldots$ *is said to be a* functional sequence.

The individual functions $f_n(x)$ are called *members* or *elements* of the sequence, and $\{x\}$ is its *domain of definition* or simply domain.

The symbol $\{f_n(x)\}$ will be used to designate a functional sequence.

The formally written sum

$$\sum_{n=1}^{\infty} u_n(x) = u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots \tag{1.1}$$

of an infinite number of elements of a functional sequence $\{u_n(x)\}$ will be called a *functional series.*

The terms $u_n(x)$ of that series are functions defined on some set $\{x\}$.

The set $\{x\}$ is called the *domain of definition*, or domain, of the functional series (1.1).

As in the case of the number series, the sum of the first n terms of (1.1) are called the nth *partial sum* of that series.

It should be stressed that the *study of functional series is perfectly equivalent to the study of functional sequences*, for to every functional series (1.1) uniquely corresponds a functional sequence

$$S_1(x),\ S_2(x),\ \ldots,\ S_n(x),\ \ldots \tag{1.2}$$

* In particular, by $\{x\}$ we may imply both the set of points of a straight line and the set of points $x = (x_1, x_2, \ldots, x_m)$ of a Euclidean space E^m.

of its partial sums, and, conversely, to every functional sequence (1.2) uniquely corresponds a functional series (1.1) with terms

$$u_1(x) = S_1(x),\ u_n(x) = S_n(x) - S_{n-1}(x) \quad \text{for} \quad n \geqslant 2$$

for which the sequence (1.2) is a sequence of partial sums.

Here are examples of functional sequences and series.

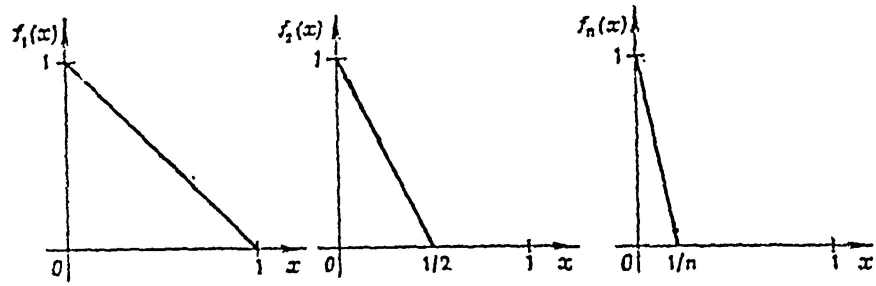


Fig. 1.1

Example 1. Consider a sequence of functions $\{f_n(x)\}$ each defined on the closed interval $0 \leqslant x \leqslant 1$ and having the form

$$f_n(x) = \begin{cases} (1-nx) & \text{when } 0 \leqslant x \leqslant 1/n, \\ 0 & \text{when } 1/n \leqslant x \leqslant 1. \end{cases} \tag{1.3}$$

Figure 1.1 gives the graphs of the functions $f_1(x)$, $f_2(x)$ and $f_n(x)$.

Example 2. As an example of a functional series consider the following power series in x:

$$1 + \sum_{k=1}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots \tag{1.4}$$

Notice that the $(n+1)$th partial sum of (1.4) differs from the Maclaurin expansion of e^x only by the remainder term $R_{n+1}(x)$.

1.1.2. Convergence of a functional sequence at a point and on a set. Suppose a functional sequence (or series) is defined on a set $\{x\}$ Fix an arbitrary point x_0 of $\{x\}$ and consider all the elements or the sequence (or the terms of the series) at x_0. We obtain a number sequence (or series).

If this number sequence (or series) converges, the given functional sequence (or series) is said *to converge at* x_0.

The set of all points x_0 at which a given functional sequence (or series) converges is called the *domain of convergence* of that sequence (or that series).

At various particular cases the domain of convergence may either coincide with the domain of definition or form a part of the domain or be an empty set.

Corresponding examples are given below.

Suppose that $\{f_n(x)\}$ has $\{x\}$ as the domain of convergence. The collection of all limits taken for all the values of x from the set $\{x\}$ forms a well-defined function $f(x)$ also given on $\{x\}$.

This function is called the *limit function* of the sequence $\{f_n(x)\}$.

Quite similarly, if the functional series (1.1) converges on some set $\{x\}$, then on that set a function $S(x)$ is defined which is the limit function of the sequence of its partial sums and is called the sum of that series.

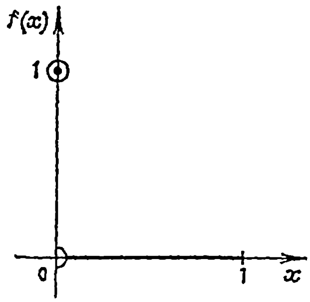


Fig. 1.2

The sequence (1.3) of Example 1 above *converges on the whole of the closed interval* $0 \leqslant x \leqslant 1$.

Indeed, $f_n(0) = 1$ for all integers n, i.e. at the point $x=0$ the sequence (1.3) converges to unity.

But if we choose any x in the half-open interval $0 < x \leqslant 1$, then all $f_n(x)$ beginning with some integer (dependent of course on x) will be zero. At any point x of $0 < x \leqslant 1$ therefore (1.3) converges to zero.

So, (1.3) converges on the entire interval $0 \leqslant x \leqslant 1$ to a limit function $f(x)$ having the form

$$f(x) = \begin{cases} 1 \text{ when } x=0 \\ 0 \text{ when } 0<x\leqslant 1. \end{cases}$$

The graph of that limit function is given in Fig. 1.2.

We stress that $f(x)$ is not continuous on the interval $0 \leqslant x \leqslant 1$ (it is discontinuous at $x = 0$).

Now we turn to the functional series (1.4) of Example 2.

That series converges at any point x of an infinite straight line and its sum equals e^x. For the proof the reader is referred to Chapter 13 of [1] (see Example 3 in Section 13.1.1)*.

1.1.3. Uniform convergence on a set. Suppose that a sequence

$$f_1(x),\ f_2(x),\ \ldots,\ f_n(x),\ \ldots \tag{1.5}$$

converges on a set $\{x\}$ to a limit function $f(x)$.

Definition 1. The sequence (1.5) *is said to converge to the function* $f(x)$ **uniformly on the set** $\{x\}$ *if for any* $\varepsilon > 0$ *we can find an integer* $N(\varepsilon)$ *such that given* $n \geqslant N(\varepsilon)$, *for all* x *of* $\{x\}$ *we have***

$$|f_n(x) - f(x)| < \varepsilon. \tag{1.6}$$

* This proof, however, immediately follows from the Maclaurin formula for e^x and from the fact that the remainder in that formula tends to zero for all x.

** If by $\{x\}$ we mean a set of points $x = (x_1, \ldots, x_m)$ of a space E^m, then we obtain a definition of uniform convergence of a sequence $f_n(x) = f_n(x_1, x_2, \ldots, x_m)$ of functions of m variables.

Remark 1. What is very essential to this definition is that N depends only on ε and is independent of x. Thus for any $\varepsilon > 0$ we can find a universal integer $N(\varepsilon)$ beginning with which inequality (1.6) holds at once for all x of $\{x\}$.

Remark 2. *The convergence of $\{f_n(x)\}$ on $\{x\}$ does not at all imply its uniform convergence on $\{x\}$.* Thus the sequence (1.3) of Example 1 above converges on the entire interval $[0, 1]$ (that was established above).

We now prove that that sequence *does not uniformly converge on* $[0, 1]$. Consider a sequence of points $x_n = 1/2n$ $(n = 1, 2, \ldots)$ belonging to $[0, 1]$. At each of the points (i.e. for every n) the relations $f_n(x_n) = 1/2$, $f(x_n) = 0$ hold. Thus for any n

$$|f_n(x_n) - f(x_n)| = 1/2,$$

i.e. when $\varepsilon \leqslant 1/2$, inequality (1.6) cannot hold for all points x of $[0, 1]$ at once, whatever n we have.

Remark 3. Note that the uniform convergence of $\{f_n(x)\}$ on $\{x\}$ to $f(x)$ is equivalent to the convergence of a number sequence $\{\varepsilon_n\}$ whose elements ε_n are the suprema of the function $|f_n(x) - f(x)|$ on $\{x\}$.

Remark 4. It is immediate from Definition 1 that if $\{f_n(x)\}$ uniformly converges to $f(x)$ on the whole of $\{x\}$, then it does so on any portion of $\{x\}$.

Now we give an example of a functional sequence uniformly converging on some set $\{x\}$. Consider again the sequence (1.3), this time not on the whole of $[0, 1]$, but on a closed interval $[\delta, 1]$, where δ is a fixed number in the interval $0 < \delta < 1$. For any such δ we can find an integer beginning with which all the elements $f_n(x)$ are zero on $[\delta, 1]$. Since the limit function $f(x)$ is also zero on $[\delta, 1]$, everywhere on $[\delta, 1]$ we have $|f_n(x) - f(x)| < \varepsilon$ for any $\varepsilon > 0$ beginning with the indicated integer. This proves the uniform convergence of (1.3) on $[\delta, 1]$.

Definition 2. A functional series is said to be uniformly convergent *on a set $\{x\}$ to its sum $S(x)$ if the sequence $\{S_n(x)\}$ of its partial sums converges uniformly on $\{x\}$ to the limit function $S(x)$.*

Prove on your own that the functional series (1.4) of Example 2 above converges to its sum e^x uniformly on every interval $-r \leqslant x \leqslant$ $\leqslant r$, where r is any fixed positive number*.

---

* To prove this it suffices to evaluate the remainder $R_{n+1}(x)$ in the Maclaurin formula for e^x. This remainder, which is the difference between e^x and the $(n+1)$th partial sum of (1.4), satisfies for all x in the interval $-r \leqslant x \leqslant r$ at once the inequality

$$|R_{n+1}(x)| \leqslant \frac{r^{n+1}}{(n+1)!} e^r$$

see formula (8.62) in [1]).

1.1.4. Cauchy criterion. The following two *main* theorems hold.

Theorem 1.1. For a functional sequence $\{f_n(x)\}$ to converge uniformly on a set $\{x\}$ to some limit function it is necessary and sufficient that for any $\varepsilon > 0$ we should be able to find $N(\varepsilon)$ such that

$$|f_{n+p}(x) - f_n(x)| < \varepsilon \tag{1.7}$$

for all $n \geqslant N(\varepsilon)$, all natural p $(p = 1, 2, \ldots)$ and all x of $\{x\}$.

Theorem 1.2. For a functional series

$$\sum_{k=1}^{\infty} u_k(x) \tag{1.8}$$

to converge uniformly on a set $\{x\}$ to a certain sum it is necessary and sufficient that for any $\varepsilon > 0$ we should be able to find $N(\varepsilon)$ such that

$$\left|\sum_{k=n+1}^{n+p} u_k(x)\right| < \varepsilon \tag{1.9}$$

for all $n \geqslant N(\varepsilon)$, all natural p and all x of $\{x\}$.

Theorem 1.2 is a consequence of Theorem 1.1: it is sufficient to notice that under the modulus sign on the left of (1.9) we have the difference $S_{n+p}(x) - S_n(x)$ of the partial sums of (1.8).

Proof of Theorem 1.1. (1) **Necessity.** Let $\{f_n(x)\}$ converge uniformly on $\{x\}$ to some limit function $f(x)$. Choose an arbitrary $\varepsilon > 0$. For the positive number $\varepsilon/2$ we can find N such that for all $n \geqslant N$ and for all x of $\{x\}$

$$|f_n(x) - f(x)| < \varepsilon/2. \tag{1.10}$$

If p is any natural number, then for $n \geqslant N$ and for all x of $\{x\}$ all the more so

$$|f_{n+p}(x) - f(x)| < \varepsilon/2 \tag{1.11}$$

Since the modulus of a sum is not greater than the sum of moduli, by (1.10) and (1.11) we get

$$|f_{n+p}(x) - f_n(x)| = |[f_{n+p}(x) - f(x)] + [f(x) - f_n(x)]| \leqslant$$
$$\leqslant |f_{n+p}(x) - f(x)| + |f(x) - f_n(x)| < \varepsilon$$

for all $n \geqslant N$, all natural p and all x of $\{x\}$). Necessity is proved.

(2) **Sufficiency.** Inequality (1.7), together with the Cauchy criterion for the number sequence, implies the convergence of $\{f_n(x)\}$ for any given x of $\{x\}$ and the existence of a limit function $f(x)$.

Since (1.7) holds *for any natural p*, by proceeding in this inequality to the limit as $p \to \infty$ (see Theorem 3.13 in [1]) we find that for all $n \geqslant N$ and for all x of $\{x\}$

$$|f(x) - f_n(x)| \leqslant \varepsilon.$$

By the arbitrariness of $\varepsilon > 0$, sufficiency is proved.

1.1.5. Sufficient tests for uniform convergence. Depending on convenience we shall formulate the tests for uniform convergence either in terms of sequences or in terms of series*.

We introduce a new concept to formulate the first test.

Definition. A sequence of functions $\{f_n(x)\}$ *is said* to be uniformly bounded on a set $\{x\}$, *if there is a real number* A *such that for all* x *of* $\{x\}$ *and for all* n *we have* $|f_n(x)| \leqslant A$.

Theorem 1.3 (Abel-Dirichlet test). Let

$$\sum_{k=1}^{\infty} u_k(x) \cdot v_k(x)$$

be a functional series. It converges uniformly on a set $\{x\}$, *if the following two conditions hold:*

(i) *the sequence* $\{v_k(x)\}$ *is nonincreasing on* $\{x\}$ *and uniformly converges on* $\{x\}$ *to zero;*

(ii) *the series* $\sum\limits_{k=1}^{\infty} u_k(x)$ *has a sequence of partial sums uniformly bounded on* $\{x\}$.

The proof coincides almost textually with that of the corresponding test for the convergence of number series (see Section 13.5.2 in [1]). The reader should carry it out on his own.

Example 1. As an illustration, we discuss uniform convergence of the series

$$\sum_{k=1}^{\infty} \frac{\sin kx}{k}. \tag{1.12}$$

Since $\{1/k\}$ (for all x) is not increasing and tends uniformly to zero, by the Abel-Dirichlet test, (1.12) converges uniformly on any set on which

$$\sum_{k=1}^{\infty} \sin kx \tag{1.13}$$

has a uniformly bounded sequence of partial sums. We compute and evaluate the nth partial sum $S_n(x)$ of (1.13).

Summing the identity

$$2\sin\frac{x}{2}\cdot\sin kx = \cos\left(k-\frac{1}{2}\right)x - \cos\left(k+\frac{1}{2}\right)x$$

over all k from 1 to n, we get

$$2\sin\frac{x}{2}\cdot S_n(x) = \cos\frac{x}{2} - \cos\left(n+\frac{1}{2}\right)x.$$

* By what was said in Section 1.1.1 both formulations are equivalent.

From this

$$S_n(x)=\frac{\cos\frac{x}{2}-\cos\left(n+\frac{1}{2}\right)x}{2\sin\frac{x}{2}}.$$

Therefore for all n

$$|S_n(x)|\leqslant\frac{1}{\left|\sin\frac{x}{2}\right|}. \tag{1.14}$$

From inequality (1.14) it is evident that the sequence $\{S_r(x)\}$ of partial sums of (1.13) is uniformly bounded on any fixed interval containing no points $x_m=2\pi m$ $(m=0, \pm 1, \pm 2, \ldots)$, for on any such interval $\left|\sin\frac{x}{2}\right|$ has a *positive* infimum.

Thus we have proved that (1.12) converges uniformly on any interval containing no points $x_m=2\pi m$, where $m=0, \pm1, \pm2, \ldots$.

Theorem 1.4. (***Weierstrass test***). *If a functional series*

$$\sum_{k=1}^{\infty} u_k(x) \tag{1.15}$$

is defined on a set $\{x\}$ and if there is a convergent number series $\sum_{k=1}^{\infty} c_k$ such that for all x of a set $\{x\}$ and any integer k

$$|u_k(x)|\leqslant c_k, \tag{1.16}$$

then (1.15) *converges uniformly on $\{x\}$.*

Concise statement: *a functional series converges uniformly on a given set if it can be majorized on that set by a convergent number series.*

Proof. By the Cauchy criterion for the number series $\sum_{k=1}^{\infty} c_k$, for any $\varepsilon>0$ there is $N(\varepsilon)$ such that for all $n\geqslant N(\varepsilon)$ and for any natural p

$$\sum_{k=n+1}^{n+p} c_k<\varepsilon. \tag{1.17}$$

From (1.16) and (1.17) and from the fact that the modulus of a sum is not greater than the sum of moduli we get

$$\left|\sum_{k=n+1}^{n+p} u_k(x)\right|<\varepsilon$$

(for all $n\geqslant N(\varepsilon)$, all natural p and all x of $\{x\}$).

By the Cauchy criterion, (1.15) converges uniformly on $\{x\}$. Thus the theorem is proved.

Example 2. The series

$$\sum_{k=1}^{\infty} \frac{\sin kx}{k^{1+\delta}}, \quad \text{where } \delta > 0,$$

converges uniformly on the whole of an infinite straight line, for on the entire straight line

$$\left| \frac{\sin kx}{k^{1+\delta}} \right| \leqslant \frac{1}{k^{1+\delta}},$$

and the number series $\sum_{k=1}^{\infty} \frac{1}{k^{1+\delta}}$, with $\delta > 0$, converges (see Chapter 13 in [1]).

Remark 1. *The Weierstrass test is not necessary.*

Indeed, it was established above that (1.12) converges uniformly on any interval containing no points $x_m = 2\pi m$ ($m = 0, \pm 1, \pm 2, \ldots$ ...). In particular, (1.12) converges uniformly on $[\pi/2, 3\pi/2]$. On this interval, however, the absolute value of the kth term of (1.12), $\frac{|\sin kx|}{k}$, has a supremum equal to $1/k$, i. e. the majorant $\sum_{k=1}^{\infty} \frac{1}{k}$ is known to be a divergent harmonic series.

Theorem 1.5 (Dini test). Let $\{f_n(x)\}$ be a sequence not decreasing (or not increasing) at each point of a close interval $[a, b]$ and converging on that interval to a limit function $f(x)$. Then, if all the elements of the sequence $f_n(x)$ and the limit function $f(x)$ are continuous on $[a, b]$, the convergence of $\{f_n(x)\}$ is uniform on $[a, b]$.*

Proof. For definiteness suppose that $\{f_n(x)\}$ is not decreasing on $[a, b]$ (the case of nonincreasing sequence can be reduced to this case by multiplying all the elements of the sequence by -1).

Put

$$r_n(x) = f(x) - f_n(x).$$

The sequence $\{r_n(x)\}$ has the following properties:

(1) all $r_n(x)$ are nonnegative and continuous on $[a, b]$;

(2) $\{r_n(x)\}$ is nonincreasing on $[a, b]$;

(3) at each point x of $[a, b]$ there is a limit $\lim\limits_{n\to\infty} r_n(x) = 0$.

We want to prove that $\{r_n(x)\}$ converges to zero uniformly on $[a, b]$. It suffices to prove that given any $\varepsilon > 0$, there is *at least one* n such that $r_n(x) < \varepsilon$ *at once for all* x *of* $[a, b]$ (then by property (2) above $r_n(x) < \varepsilon$ for all subsequent integers too).

* Ulisse Dini (1845-1918) is an Italian mathematician.

Suppose given some $\varepsilon > 0$ we can find no n such that $r_n(x) < \varepsilon$ at once for all x in $[a, b]$. Then for any n there is a point x_n in $[a, b]$ such that

$$r_n(x_n) \geqslant \varepsilon. \tag{1.18}$$

By the Bolzano-Weierstrass theorem, we may choose a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ converging to some point x_0 in $[a, b]$ (see Section 3.4 in [1]).

All $r_m(x)$ (given any m) are continuous at x_0. Therefore for any m

$$\lim_{k\to\infty} r_m(x_{n_k}) = r_m(x_0). \tag{1.19}$$

On the other hand, on choosing for any given m some n_k greater than m we obtain (by property (2) above)

$$r_m(x_{n_k}) \geqslant r_{n_k}(x_{n_k}).$$

Comparing the last inequality with (1.18) we have

$$r_m(x_{n_k}) \geqslant \varepsilon \tag{1.20}$$

(for any given m and any n_k greater than m). Finally, comparing (1.19) and (1.20) yields

$$r_m(x_0) \geqslant \varepsilon$$

(for any m).

The last inequality contradicts the convergence of $\{r_n(x)\}$ to zero at x_0. This contradiction proves the theorem.

Remark 2. Essential in the Dini theorem is the condition of *monotonicity* of $\{f_n(x)\}$ on $[a, b]$, for a sequence nonmonotonic on $[a, b]$ of functions continuous on $[a, b]$ may converge at each point of $[a, b]$ to a function $f(x)$ continuous on $[a, b]$, but without converging to it uniformly on $[a, b]$.

A sequence of functions $f_n(x)$ equal to $\sin nx$ for $0 \leqslant x \leqslant \pi/n$ and to zero for $\pi/n < x \leqslant \pi$ $(n = 1, 2, \ldots)$ may serve as an example. It converges to $f(x) \equiv 0$ at each point of $[0, \pi]$ but is not uniformly convergent on $[0, \pi]$, for $|f_n(x_n) - f(x_n)| = 1$, given $x_n = \pi/2n$ for all n.

Remark 3. We formulate the Dini theorem in terms of series: *if all terms of a series are continuous and nonnegative on a closed interval* $[a, b]$ *and the sum of the series is also continuous on* $[a, b]$, *then that series uniformly converges to its sum on* $[a, b]$.

Remark 4. The Dini theorem and its proof remain valid if we take any bounded closed set $\{x\}$ instead of $[a, b]$ in the theorem. It is customary to call such a set a *compact* set.

Example 3. The sequence $\{x^n\}$ uniformly converges to zero on a closed interval $\left[0, \frac{1}{2}\right]$.

Indeed, (1) for any x in $\left[0, \frac{1}{2}\right]$ the sequence converges to zero; (2) all the functions x^n and the limit function zero are continuous on $\left[0, \frac{1}{2}\right]$; (3) $\{x^n\}$ is not increasing on $\left[0, \frac{1}{2}\right]$.

Thus all the hypotheses of the Dini theorem are satisfied.

1.1.6. Proceeding to the limit term by term. Continuity of the sum of a series and of the limit function of a sequence. Consider a point a on an infinite straight line. Let $\{x\}$ be a set that may not contain a but such that there are points of the set $\{x\}$ in any ε-neighbourhood of a*.

The following statement holds.

Theorem 1.6. Let a functional series

$$\sum_{k=1}^{\infty} u_k(x) \tag{1.15}$$

converge uniformly on a set $\{x\}$ to a sum $S(x)$. Further let all terms of that series have at a point a a limiting value

$$\lim_{x\to a} u_k(x) = b_k.$$

Then the function $S(x)$ also has a limiting value at a, with

$$\lim_{x\to a} S(x) = \sum_{k=1}^{\infty} \lim_{x\to a} u_k(x) = \sum_{k=1}^{\infty} b_k, \tag{1.21}$$

i.e. the limit sign lim *and the summation sign* $\sum$ *may be interchanged (or, as we say, we can proceed to the limit* term by term).

Proof. We first prove that the number series $\sum_{k=1}^{\infty} b_k$ converges. By the Cauchy criterion applied to (1.15), given any $\varepsilon > 0$, we can find $N(\varepsilon)$ such that

$$|u_{n+1}(x) + u_{n+2}(x) + \ldots + u_{n+p}(x)| < \varepsilon \tag{1.22}$$

for all $n \geqslant N(\varepsilon)$, all natural p and all x of $\{x\}$.

Proceeding in (1.22) to the limit $x \to a$** we get

$$|b_{n+1} + b_{n+2} + \ldots + b_{n+p}| \leqslant \varepsilon < 2\varepsilon$$

(for all $n \geqslant N(\varepsilon)$ and all natural p).

The Cauchy criterion holds for $\sum_{k=1}^{\infty} b_k$ therefore, and the series converges.

---

* In other words, a is the limit point of $\{x\}$.

** To do this we can use some sequence of points $\{x_n\}$ converging to a.

Now we evaluate the difference $S(x)-\sum_{k=1}^{\infty} b_k$ for all values of x in a small neighbourhood of a. Since $S(x)=\sum_{k=1}^{\infty} u_k(x)$ for all points of $\{x\}$, given any n we have

$$S(x)-\sum_{k=1}^{\infty} b_k \equiv \left[\sum_{k=1}^{n} u_k(x)-\sum_{k=1}^{n} b_k\right]+\sum_{k=n+1}^{\infty} u_k(x)-\sum_{k=n+1}^{\infty} b_k.$$

From this we obtain for all x of $\{x\}$

$$\left|S(x)-\sum_{k=1}^{\infty} b_k\right| \leqslant \left|\sum_{k=1}^{n} u_k(x)-\sum_{k=1}^{n} b_k\right|+\left|\sum_{k=n+1}^{\infty} u_k(x)\right|+$$
$$+\left|\sum_{k=n+1}^{\infty} b_k\right|. \tag{1.23}$$

Take $\varepsilon>0$. Since the series $\sum_{k=1}^{\infty} b_k$ converges and since (1.15) converges uniformly on $\{x\}$, for given ε we can find n such that for all x of $\{x\}$

$$\left|\sum_{k=n+1}^{\infty} b_k\right|<\frac{\varepsilon}{3}, \quad \left|\sum_{k=n+1}^{\infty} u_k(x)\right|<\frac{\varepsilon}{3}. \tag{1.24}$$

Since the limit of a finite sum is equal to the sum of the limits of summands, for given $\varepsilon>0$ and the chosen n there is $\delta>0$ such that

$$\left|\sum_{k=1}^{n} u_k(x)-\sum_{k=1}^{n} b_k\right|<\frac{\varepsilon}{3} \tag{1.25}$$

for all x of $\{x\}$ satisfying the condition $0<|x-a|<\delta$.

Putting (1.24) and (1.25) in the right-hand side of (1.23) we finally get

$$\left|S(x)-\sum_{k=1}^{\infty} b_k\right|<\varepsilon$$

for all the points x of $\{x\}$ satisfying the condition $0<|x-a|<\delta$. This proves that the function $S(x)$ has a limiting value at $x=a$ and that equation (1.21) holds. Thus the theorem is proved.

Now we formulate Theorem 1.6 in terms of functional sequences. *If a functional sequence $\{f_n(x)\}$ converges uniformly to a limit function $f(x)$ on a set $\{x\}$ and if each element of the sequence has a limiting value at a point a, then $f(x)$ also has a limiting value at a,*

with

$$\lim_{x\to a} f(x) = \lim_{n\to\infty}(\lim_{x\to a} f_n(x)) = \lim_{n\to\infty}(\lim_{x\to a} f(x)),$$

i.e. the sign of the limit of a sequence $\lim\limits_{n\to\infty}$ *and the sign of the limiting value of a function* $\lim\limits_{x\to a}$ *may be interchanged (or as we say we can proceed to the limit as* $x \to a$ element by element).

Remark to Theorem 1.6. If under the hypotheses of Theorem 1.6 we in addition require that a should belong to $\{x\}$ and that all terms $u_l(x)$ of (1.15) should be continuous at a (or respectively continuous at a on the right or the left), then the sum $S(x)$ of (1.15) is also continuous at a (or respectively continuous at a on the right or the left).

Indeed, in this case $b_k = u_k(a)$ and equation (1.21) becomes

$$\lim_{x\to a} S(x) = \sum_{k=1}^{\infty} u_k(a) = S(a),$$

which just implies the continuity of the function $S(x)$ at a (or, if x tends to a from one side, the continuity of $S(x)$ at a respectively on the right or on the left).

Applying the above remark to each point of some closed interval $[a, b]$ we arrive at the following main theorem.

Theorem 1.7. If all terms of a functional series (all elements of a functional sequence) are continuous on a closed interval $[a, b]$ *and if that series (that sequence) uniformly converges on* $[a, b]$, *then the sum of the series (the limit function of the sequence) is also continuous on* $[a, b]$.

Remarks to Theorem 1.7. (1) In Theorem 1.7, instead of a closed interval $[a, b]$ we may take an open interval, a half-open interval, a half-line, an infinite straight line or any set $\{x\}$ dense in itself in general. (2) Essential in Theorem 1.7 is the requirement on *uniform* convergence, for a nonuniformly convergent sequence of continuous functions may converge to a discontinuous function (see Example (1.3) in Sections 1.1.1 and 1.1.2).

Concluding remark. All the theorems of this section are true for sequences of functions given on a set $\{x\}$ of E^m.

1.2. TERM-BY-TERM INTEGRATION AND TERM-BY-TERM DIFFERENTIATION OF FUNCTIONAL SEQUENCES AND SERIES

1.2.1. Term-by-term integration. The following *main* theorem holds.

Theorem 1.8. If a functional sequence $\{f_n(x)\}$ *uniformly converges to a limit function* $f(x)$ *on a closed interval* $[a, b]$ *and if every function* $f_n(x)$ *is integrable on* $[a, b]$, *then* $f(x)$ *is also integrable on* $[a, b]$, *it*

being possible to integrate $\{f_n(x)\}$ *over* $[a, b]$ element by element, *i.e.*

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx$$

exists and is equal to $\int_a^b f(x)\,dx$.

Proof. Choose $\varepsilon > 0$. Then, by the uniform convergence of $\{f_n(x)\}$ to $f(x)$, there is $N(\varepsilon)$ for $\varepsilon > 0$ such that for all $n \geqslant N(\varepsilon)$ and for all x in $[a, b]$

$$|f_n(x) - f(x)| < \frac{\varepsilon}{2(b-a)}. \tag{1.26}$$

If we prove that $f(x)$ is integrable on $[a, b]$, then, using the ordinary evaluations of integrals* and inequality (1.26), we get

$$\left|\int_a^b f_n(x)\,dx - \int_a^b f(x)\,dx\right| = \left|\int_a^b [f_n(x) - f(x)]\,dx\right| \leqslant$$

$$\leqslant \int_a^b |f_n(x) - f(x)|\,dx \leqslant \frac{\varepsilon}{2(b-a)} \int_a^b dx = \frac{\varepsilon}{2} < \varepsilon$$

(for all $n \geqslant N(\varepsilon)$).

This proves that the limit $\lim\limits_{n\to\infty} \int_a^b f_n(x)\,dx$ exists and is equal to $\int_a^b f(x)\,dx$, and it remains for us to prove that $f(x)$ is integrable on $[a, b]$.

On dividing $[a, b]$ by means of *arbitrary* points $a = x_0 < x_1 < \ldots \ldots < x_m = b$ into m closed intervals $[x_{k-1}, x_k]$ $(k = 1, 2, \ldots, m)$ we agree to denote by $\omega_k(f)$ (respectively by $\omega_k(f_n)$) oscillation on the kth closed interval $[x_{k-1}, x_k]$ of $f(x)$ (respectively of $f_n(x)$)**

* We mean the following evaluations of integrals established in Section 10.6 of [1]: (1) if a function $F(x)$ is integrable on $[a, b]$, so is $|F(x)|$ with $\left|\int_a^b F(x)\,dx\right| \leqslant \int_a^b \left|F(x)\right| dx$; (2) if $f(x)$ and $g(x)$ are both integrable on $[a, b]$ and everywhere on it $f(x) \leqslant g(x)$, then $\int_a^b f(x)\,dx \leqslant \int_a^b g(x)\,dx$.

** Recall that the *oscillation* of a function on a given closed interval is the difference between the supremum and infimum of the function on the given interval.

We want to see that given any $\varepsilon > 0$ and any $k = 1, 2, \ldots, m$, we can find a sufficiently large n for which

$$\omega_k(f) \leqslant \omega_k(f_n) + \frac{\varepsilon}{b-a}. \tag{1.27}$$

Indeed, whatever x' and x'' of $[x_{k-1}, x_k]$, we have

$$|f(x') - f(x'')| \leqslant |f(x') - f_n(x')| + |f_n(x') - f_n(x'')| + $$
$$+ |f_n(x'') - f(x'')|. \tag{1.28}$$

By the uniform convergence of $\{f_n(x)\}$ to $f(x)$, given any $\varepsilon > 0$, there is n such that for any x in $[a, b]$ inequality (1.26) holds. Thus for this n

$$|f(x') - f_n(x')| + |f_n(x'') - f(x'')| < \frac{\varepsilon}{b-a}$$

and therefore, by (1.28),

$$|f(x') - f(x'')| \leqslant |f_n(x') - f_n(x'')| + \frac{\varepsilon}{b-a}.$$

From the last inequality and from the arbitrariness of points x' and x'' it immediately follows that inequality (1.27) holds for the chosen integer n.

Denote now, for the arbitrary subdivision of $[a, b]$ we have assumed, the upper and lower sums of $f(x)$ by S and s and the upper and lower sums of $f_n(x)$ by S_n and s_n.

Multiplying inequality (1.27) by the length of the kth closed interval Δx_k and then summing it over all $k = 1, 2, \ldots, m$ we get

$$S - s \leqslant S_n - s_n + \varepsilon. \tag{1.29}$$

Inequality (1.29) is established by us for an arbitrary subdivision of $[a, b]$. By the integrability of the function $f_n(x)$ on $[a, b]$ we can find a subdivision of $[a, b]$ for which $S_n - s_n < \varepsilon$* and therefore, by (1.29), $S - s < 2\varepsilon$.

Since ε is an arbitrary positive number, the last inequality proves that $f(x)$ is integrable on $[a, b]$**. Thus the theorem is proved.

Now we formulate Theorem 1.8 in terms of functional series:

If a functional series (1.15) converges to its sum $S(x)$ uniformly on a closed interval $[a, b]$ and if each term of that series $u_k(x)$ is a function integrable on $[a, b]$, then $S(x)$ is also integrable on $[a, b]$, it being

* By Theorem 10.1 of [1].

** By Theorem 10.1 of [1], the existence for an arbitrary $\varepsilon > 0$ of a subdivision of a closed interval for which $S - s < 2\varepsilon$ is a necessary and sufficient condition for any function *bounded* on the interval to be integrable. The boundedness of $f(x)$ on $[a, b]$ immediately follows from (1.26) and from the boundedness of the function $f_n(x)$ integrable on $[a, b]$.

possible to integrate (1.15) *over* $[a, b]$ **term by term**, *i.e. the series*

$$\sum_{k=1}^{\infty} \int_a^b u_k(x)\,dx$$

converges and has as its sum $\int_a^b S(x)\,dx$.

Remark. In courses of mathematical analysis Theorem 1.8 tends to be proved under the more stringent assumption that every function $f_n(x)$ is not only integrable but is continuous on $[a, b]$ as well. With this additional assumption the above proof becomes simpler, for it is sufficient to refer to Theorem 1.7 to prove that the limit function $f(x)$ is integrable on $[a, b]$.

1.2.2. Term-by-term differentiation. Now we prove the following *main* theorem.

Theorem 1.9. Let every function $f_n(x)$ *have a derivative* $f_n'(x)$ *on a closed interval* $[a, b]$*, *the sequence of derivatives* $\{f_n'(x)\}$ *converging uniformly on* $[a, b]$ *and the sequence* $\{f_n(x)\}$ *itself converging at least at one point* x_0 *of* $[a, b]$. *Then* $\{f_n(x)\}$ *uniformly converges to some limit function* $f(x)$ *on the entire interval* $[a, b]$, $\{f_n(x)\}$ *being differentiable on* $[a, b]$ **element by element**, *i.e. everywhere on* $[a, b]$ $f(x)$ *has the derivative* $f'(x)$, *which is the limit function of* $\{f_n'(x)\}$.

Proof. We first prove that $\{f_n(x)\}$ *uniformly converges on* $[a, b]$. From the convergence of the number sequence $\{f_n(x_0)\}$ and from the uniform convergence on $[a, b]$ of $\{f_n'(x)\}$ we deduce that given an arbitrary $\varepsilon > 0$ we can find $N(\varepsilon)$ such that

$$|f_{n+p}(x_0) - f_n(x_0)| < \frac{\varepsilon}{2}, \quad |f_{n+p}'(x) - f_n'(x)| < \frac{\varepsilon}{2(b-a)} \quad (1.30)$$

for all $n \geqslant N(\varepsilon)$, all natural p and (this applies to the second of the inequalities (1.30)) all x of $[a, b]$.

Let x be an *arbitrary* point in $[a, b]$. Given any fixed n and p, all the hypotheses of the Lagrange theorem hold on $[x_0, x]$ for the function $[f_{n+p}(t) - f_n(t)]$ (see Theorem 8.12 in [1]). By that theorem there is a point ξ between x and x_0 such that

$$[f_{n+p}(x) - f_n(x)] - [f_{n+p}(x_0) - f_n(x_0)] =$$
$$= [f_{n+p}'(\xi) - f_n'(\xi)](x - x_0).$$

From this last equation and from (1.30), on taking into account the fact that $|x - x_0| \leqslant b - a$, we get

$$|f_{n+p}(x) - f_n(x)| < \varepsilon$$

(for any x in $[a, b]$, any $n \geqslant N(\varepsilon)$ and any natural p).

* Here and in what follows by "$f(x)$ has a derivative on a closed interval $[a, b]$" we mean the existence of a derivative $f'(x)$ at any interior point of $[a, b]$, the right-hand derivative $f'(a + 0)$ at a and the left-hand derivative $f'(b - 0)$ at b.

But this just means that $\{f_n(x)\}$ converges uniformly on $[a, b]$ to some limit function $f(x)$*.

It remains to prove that $f(x)$ *has a derivative at any point* x_0 *in* $[a, b]$ *and that that derivative is the limit function of* $\{f_n'(x)\}$.

Choose an *arbitrary* point x_0 and, correspondingly, a positive number δ such that the δ-neighbourhood of x_0 is contained entirely in $[a, b]$ (in case x_0 is an end point of $[a, b]$, by the δ-neighbourhood of x_0 we shall mean either the right-hand half-neighbourhood $[a, a+\delta)$ of the point a or respectively the left-hand half-neighbourhood $(b-\delta, b]$ of b).

Denote by $\{\Delta x\}$ a set of all numbers Δx satisfying the condition $0 < |\Delta x| < \delta$ when $a < x_0 < b$, the condition $0 < \Delta x < \delta$ when $x_0 = a$ or the condition $-\delta < \Delta x < 0$ when $x_0 = b$ and prove that the sequence of functions of the argument Δx

$$\varphi_n(\Delta x) = \frac{f_n(x_0+\Delta x) - f_n(x_0)}{\Delta x}$$

converges uniformly on the set $\{\Delta x\}$.

Given an arbitrary $\varepsilon > 0$, by the uniform convergence of $\{f_n'(x)\}$ there is $N(\varepsilon)$ such that

$$|f_{n+p}'(x) - f_n'(x)| < \varepsilon \tag{1.31}$$

for all x in $[a, b]$, all $n \geqslant N(\varepsilon)$ and all natural p.

Observing this, choose an arbitrary Δx of the set $\{\Delta x\}$ and apply the Lagrange theorem to $[f_{n+p}(t) - f_n(t)]$ (with any n and p chosen) on $[x_0, x_0+\Delta x]$. By the theorem, there is a number θ in $0 < \theta < 1$ such that

$$\varphi_{n+p}(\Delta x) - \varphi_n(\Delta x) =$$
$$= \frac{[f_{n+p}(x_0+\Delta x) - f_n(x_0+\Delta x)] - [f_{n+p}(x_0) - f_n(x_0)]}{\Delta x} =$$
$$= f_{n+p}'(x_0+\theta\Delta x) - f_n'(x_0+\theta\Delta x).$$

From the last equation and from inequality (1.31) valid for *all* the points x of $[a, b]$ we get

$$|\varphi_{n+p}(\Delta x) - \varphi_n(\Delta x)| < \varepsilon$$

for any Δx of $\{\Delta x\}$, any $n \geqslant N(\varepsilon)$ and any natural p. Thus $\{\varphi_n(\Delta x)\}$ converges uniformly on $\{\Delta x\}$ (by the Cauchy criterion). But this allows Theorem 1.6 on proceeding to the limit term by term to be applied to $\{\varphi_n(\Delta x)\}$ at $\Delta x = 0$. According to Theorem 1.6**,

$$\frac{f(x_0+\Delta x) - f(x_0)}{\Delta x},$$

---

* By the Cauchy criterion, i.e. by Theorem 1.1.

** The formulation of Theorem 1.6 in terms of functional sequences is used.

the limit function $\{\varphi_n(\Delta x)\}$, has a limiting value as $\Delta x \to 0$, with

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}=\lim_{\Delta x\to 0}[\lim_{n\to\infty}\varphi_n(\Delta x)]=$$
$$=\lim_{n\to\infty}[\lim_{\Delta x\to 0}\varphi_n(\Delta x)]=\lim_{n\to\infty}\left[\lim_{\Delta x\to 0}\frac{f_n(x_0+\Delta x)-f_n(x_0)}{\Delta x}\right]=$$
$$=\lim_{n\to\infty}f_n'(x_0)$$

This proves that the derivative of $f(x)$ at x_0 exists and is equal to $\lim\limits_{n\to\infty} f_n'(x_0)$. The theorem is proved.

Now we formulate Theorem 1.9 in terms of functional series.

If every function $u_k(x)$ has a derivative on a closed interval $[a, b]$ and if the series of derivatives $\sum\limits_{k=1}^{\infty} u_k'(x)$ converges uniformly on $[a, b]$ and the series $\sum\limits_{k=1}^{\infty} u_k(x)$ itself converges at least at one point of the interval $[a, b]$, then $\sum\limits_{k=1}^{\infty} u_k(x)$ converges uniformly on the entire interval $[a, b]$ to a certain sum $S(x)$, this series being differentiable on $[a, b]$ term by term, *i. e. its sum $S(x)$ has on $[a, b]$ a derivative which is the sum of the series of derivatives $\sum\limits_{k=1}^{\infty} u_k'(x)$.*

Remark 1. We stress that Theorem 1.9 assumes only that every function $f_n(x)$ has a derivative on $[a, b]$. Neither the boundedness nor, moreover, the integrability or continuity of that derivative is required. It is usual in courses of mathematical analysis to prove Theorem 1.9 under the additional hypothesis that every derivative $f_n'(x)$ on $[a, b]$ is continuous.

Remark 2. If Theorem 1.9 requires in addition that every derivative $f_n'(x)$ should be continuous on $[a, b]$, then by Theorem 1.7 the derivative of the limit function $f(x)$ is also continuous on $[a, b]$.

Remark 3. For the case of functions of m variables Theorem 1.9 assumes the following form: if every function $f_n(x)=f_n\times \times(x_1, \ldots, x_m)$ has on a bounded set $\{x\}$ of points in E^m a partial derivative $\frac{\partial f_n}{\partial x_k}$ and if the sequence $\left\{\frac{\partial f_n}{\partial x_k}\right\}$ converges uniformly on $\{x\}$ and the sequence $\{f_n(x)\}$ itself converges at each point of $\{x\}$, then $\{f_n(x)\}$ can be differentiated with respect to a variable x_k on $\{x\}$ *element by element.*

Theorem 1.9 yields the following statement.

Theorem 1.10. *If every function $f_n(x)$ has an antiderivative on $[a, b]$ and if the sequence $\{f_n(x)\}$ converges uniformly on $[a, b]$ to a limit function $f(x)$, then $f(x)$ also has an antiderivative on $[a, b]$. Moreover,*

if x_0 is any point of $[a, b]$, then the sequence of antiderivatives $\Phi_n(x)$ of the functions $f_n(x)$ satisfying the condition $\Phi_n(x_0) = 0$ converges uniformly on $[a, b]$ to the antiderivative $\Phi(x)$ of $f(x)$ satisfying the condition $\Phi(x_0) = 0$.

Proof. It suffices to notice that for the sequence of antiderivatives $\Phi_n(x)$ satisfying the condition $\Phi_n(x_0) = 0$ all the hypotheses of Theorem 1.9 hold. This ensures that the sequence $\{\Phi_n(x)\}$ converges uniformly on $[a, b]$ to the limit function $\Phi(x)$ which at each point of $[a, b]$ has a derivative equal to the limit function $f(x)$ of $\{f_n(x)\}$.

Remark 4. We stress that Theorem 1.10 requires neithert he boundedness nor, moreover, the integrability of functions $f_n(x)$ on $[a, b]$.

The last three subsections allow us to draw the following important **conclusion:** *uniform convergence does not change the class of functions having a limiting value* (Theorem 1.6), *the class of continuous functions* (Theorem 1.7), *the class of integrable functions* (Theorem 1.8), *the class of functions having an antiderivative* (Theorem 1.10) *or the class (in the case of uniform convergence of derivatives) of differentiable functions* (Theorem 1.9).

In conclusion consider an example, based on Theorem 1.9, of a function $f(x)$ whose derivative $f'(x)$ exists everywhere on a closed interval $[0, 1]$ but is discontinuous at each rational point of the interval.

Let

$$\varphi(x) = \begin{cases} x^2 \cos\frac{1}{x} & \text{when } x \neq 0, \\ 0 & \text{when } x = 0, \end{cases}$$

so that the function

$$\varphi'(x) = \begin{cases} \sin\frac{1}{2} + 2x \cdot \cos\frac{1}{x} & \text{when } x \neq 0, \\ 0 & \text{when } x = 0 \end{cases}$$

is discontinuous at $x = 0$ and continuous at all the other points. We number all the rational points of $[0, 1]$ to form the sequence $x_1, x_2, \ldots, x_k, \ldots$ (the possibility of this was proved in Section 3.4.3 of [1]) and put $u_k(x) = \frac{1}{k^2} \cdot \varphi \times$ $\times (x - x_k)$. Then every derivative $u_k'(x) = \frac{1}{k^2}\varphi'(x - x_k)$ is discontinuous at one point x_k and continuous at all the other points. Since for all x of $[0, 1]$

$$|u_k(x)| \leqslant \frac{|x - x_k|^2}{k^2} \leqslant \frac{1}{k^2}, \quad |u_k'(x)| \leqslant \frac{1 + 2|x - x_k|}{k^2} \leqslant \frac{3}{k^2},$$

both series $\sum_{k=1}^{\infty} u_k(x)$ and $\sum_{k=1}^{\infty} u_k'(x)$, are majorized by the convergent number series $3\sum_{k=1}^{\infty}\frac{1}{k^2}$ and therefore converge uniformly on $[0, 1]$. By Theorem 1.9

the sum $f(x)$ of $\sum_{k=1}^{\infty} u_k(x)$ has on $[0, 1]$ a derivative $f'(x)$ equal to the sum of $\sum_{k=1}^{\infty} u'_k(x)$ and discontinuous at every point x_k ($k = 1, 2, \ldots$).

1.2.3. Convergence in the mean. Suppose every function $f_n(x)$ ($n = 1, 2, \ldots$) and the function $f(x)$ are integrable on a closed interval $[a, b]$. Then (as is known from Chapter 10 of [1]) so is the function

$$[f_n(x) - f(x)]^2 = f_n^2(x) + f^2(x) - 2f_n(x) \cdot f(x).$$

We introduce the fundamental concept of convergence in the mean.

Definition 1. A sequence $\{f_n(x)\}$ is said to converge in the mean *to a function $f(x)$ on a closed interval $[a, b]$, if*

$$\lim_{n\to\infty} \int_a^b [f_n(x) - f(x)]^2\, dx = 0.$$

Definition 2. A functional series is said to converge in the mean *to a function $S(x)$ on a closed interval $[a, b]$ if the sequence of partial sums of that series converges in the mean to $S(x)$ on $[a, b]$.*

Remark. It follows from the definitions that if a sequence (or a series) converges in the mean to $f(x)$ on the entire interval $[a, b]$, that sequence (or that series) converges in the mean to $f(x)$ also on any closed interval $[c, d]$ contained in $[a, b]$.

Now we wish to show the relation between convergence in the mean and uniform convergence of a sequence.

We first prove that *if a sequence $\{f_n(x)\}$ converges uniformly to $f(x)$ on $[a, b]$, then $\{f_n(x)\}$ converges also in the mean to $f(x)$ on $[a, b]$.*

Take an arbitrary $\varepsilon > 0$. Given a positive number $\sqrt{\frac{\varepsilon}{2(b-a)}}$, by uniform convergence there is N such that

$$|f_n(x) - f(x)| < \sqrt{\frac{\varepsilon}{2(b-a)}} \tag{1.32}$$

for all x in $[a, b]$ and all $n \geqslant N$.

By (1.32), for all $n \geqslant N$

$$\int_a^b [f_n(x) - f(x)]^2\, dx \leqslant \frac{\varepsilon}{2(b-a)} \cdot \int_a^b dx = \frac{\varepsilon}{2} < \varepsilon,$$

i.e. $\{f_n(x)\}$ converges in the mean to $f(x)$ on $[a, b]$.

Now we shall show that *convergence of a sequence on some closed interval in the mean does not imply not only uniform convergence that interval but also conv[illegible]ce at least at one point of the*

Consider a sequence of closed intervals $I_1, I_2, \ldots$ belonging to $[0, 1]$ and having the following form:

$$I_1 = [0, 1],$$

$$I_2 = \left[0, \frac{1}{2}\right], \quad I_3 = \left[\frac{1}{2}, 1\right],$$

$$I_4 = \left[0, \frac{1}{4}\right], \quad I_5 = \left[\frac{1}{4}, \frac{1}{2}\right],$$

$$I_6 = \left[\frac{1}{2}, \frac{3}{4}\right], \quad I_7 = \left[\frac{3}{4}, 1\right],$$

. . . .

$$I_{2^n} = \left[0, \frac{1}{2^n}\right], \quad I_{2^n+1} = \left[\frac{1}{2^n}, \frac{2}{2^n}\right], \ldots,$$

$$I_{2^{n+1}-1} = \left[1 - \frac{1}{2^n}, 1\right],$$

. . . .

We define the nth element of the sequence as follows

$$f_n(x) = \begin{cases} 1 & \text{on } I_n, \\ 0 & \text{at the other points of } [0, 1]. \end{cases}$$

The sequence we have constructed *converges in the mean to* $f(x) = 0$ on $[0, 1]$.

Indeed,

$$\int_0^1 [f_n(x) - 0]^2\, dx = \int_{I_n} f_n^2(x)\, dx = \int_{I_n} dx =$$

$$= \text{length of closed interval } I_n \to 0 \text{ (as } n \to \infty).$$

At the same time the sequence *converges at no point of* $[0, 1]$.

Indeed, whatever point x_0 of $[0, 1]$ we may take, among *arbitrarily large* n we can find both such for which I_n contains a point x_0 (for these integers $f_n(x_0) = 1$) and such for which I_n does not contain x_0 (for these integers $f_n(x_0) = 0$). Thus $\{f_n(x_0)\}$ contains infinitely many elements, both equal to unity and zero, i.e. this sequence diverges.

It turns out that convergence of $\{f_n(x)\}$ to a limit function $f(x)$ on $[a, b]$ in the mean ensures that $\{f_n(x)\}$ can be integrated over $[a, b]$ element by element.

Theorem 1.11. If a sequence $\{f_n(x)\}$ *converges in the mean to a function* $f(x)$ *on a closed interval* $[a, b]$, *then that sequence can be integrated over* $[a, b]$ *element by element, i.e.*

$$\lim_{n\to\infty} \int_a^b f_n(x)\, dx$$

exists and is equal to $\int_a^b f(x)\, dx$.

We first prove the following lemma.

Lemma 1. For any functions $f(x)$ and $g(x)$ integrable on a closed interval $[a, b]$ we have the inequality

$$\left|\int_a^b f(x)\, g(x)\, dx\right| \leqslant \sqrt{\int_a^b f^2(x)\, dx \int_a^b g^2(x)\, dx} \tag{1.33}$$

called the **Cauchy-Buniakowski inequality.**

Proof of Lemma 1. Consider the following quadratic trinomial in λ:

$$\int_a^b [f(x)-\lambda g(x)]^2\, dx =$$

$$= \int_a^b f^2(x)\, dx - 2\lambda \int_a^b f(x)\, g(x)\, dx + \lambda^2 \int_a^b g^2(x)\, dx \geqslant 0.$$

Since this trinomial is *nonnegative*, it *has no distinct real roots.* But then its discriminant is nonpositive, i.e.

$$\left(\int_a^b f(x)\, g(x)\, dx\right)^2 - \int_a^b f^2(x)\, dx \int_a^b g^2(x)\, dx \leqslant 0.$$

Thus the lemma is proved.

Proof of Theorem 1.11. Using inequality (1.33) for $g(x) \equiv 1$ we have

$$\left|\int_a^b f_n(x)\, dx - \int_a^b f(x)\, dx\right| = \left|\int_a^b [f_n(x)-f(x)]\, dx\right| \leqslant$$

$$\leqslant \sqrt{\int_a^b [f_n(x)-f(x)]^2\, dx \int_a^b dx} =$$

$$= \sqrt{(b-a)\int_a^b [f_n(x)-f(x)]^2\, dx} \to 0$$

(as $n \to \infty$). The theorem is proved.

1.3. EQUICONTINUITY OF A SEQUENCE OF FUNCTIONS. THE ARZELA THEOREM

Suppose each of the functions $f_n(x)$ is defined on some closed interval $[a, b]$.

Definition. A sequence of functions $\{f_n(x)\}$ is said **to be equicontinuous on** $[a, b]$ *if given any $\varepsilon > 0$ we can find $\delta > 0$ such that*

$$|f_n(x') - f_n(x'')| < \varepsilon$$

holds for all n and for all points x' and x'' of $[a, b]$ related by

$|x' - x''| < \delta$.

Remark 1. It is immediate from the definition that if $\{f_n(x)\}$ is equicontinuous on $[a, b]$, then so is any of its subsequences.

We prove the following remarkable statement.

Theorem 1.12 (Arzela). If a sequence of functions $\{f_n(x)\}$ is equicontinuous and uniformly bounded on a closed interval $[a, b]$, then we can choose a subsequence uniformly converging on $[a, b]$.

Proof. Consider on $[a, b]$ the following sequence of points $\{x_n\}$: take as x_1 the point bisecting $[a, b]$, as x_2 and x_3 the two points quar-

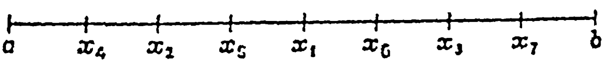


Fig. 1.3

tering, together with x_1, the interval $[a, b]$ (Fig. 1.3), as x_4, x_5, x_6, and x_7 the four points dividing, together with x_1, x_2, and x_3, the interval $[a, b]$ into eight equal parts (Fig. 1.3) and so on.

The sequence $\{x_n\}$ we have constructed has the following *property*: whatever $\delta > 0$ we take, we can find for it n_0 such that on any interval of $[a, b]$ of length δ there is at least one of the elements $x_1, x_2, \ldots, x_{n_0}$*.

Now we proceed to choose a subsequence of $\{f_n(x)\}$ uniformly converging on $[a, b]$. First consider $\{f_n(x)\}$ at x_1. We obtain a *bounded* number sequence $\{f_n(x_1)\}$; by the Bolzano-Weierstrass theorem (see Section 3.4 of [1]) we can choose a convergent (or regular) subsequence of $\{f_n(x_1)\}$ which we designate

$$f_{11}(x_1),\ f_{12}(x_1),\ \ldots,\ f_{1n}(x_1),\ \ldots$$

Then consider the sequence

$$f_{11}(x),\ f_{12}(x),\ \ldots,\ f_{1n}(x),\ \ldots$$

at x_2. By the Bolzano-Weierstrass theorem we can choose a convergent subsequence which we designate

$$f_{21}(x_2),\ f_{22}(x_2),\ \ldots,\ f_{2n}(x_2),\ \ldots$$

Thus the sequence

$$f_{21}(x),\ f_{22}(x),\ \ldots,\ f_{2n}(x),\ \ldots \tag{1.34}$$

is convergent at both x_1 and x_2.

Next consider (1.34) at x_3 and choose a convergent subsequence

$$f_{31}(x_3),\ f_{32}(x_3),\ \ldots,\ f_{3n}(x_3),\ \ldots$$

* A sequence having this property is said to be *everywhere dense* on $[a, b]$.

Continuing our arguments in a similar manner we obtain an infinite number of subsequences

$$\begin{array}{l} f_{11}(x),\ f_{12}(x),\ f_{13}(x),\ \ldots,\ f_{1n}(x),\ \ldots \\ f_{21}(x),\ f_{22}(x),\ f_{23}(x),\ \ldots,\ f_{2n}(x),\ \ldots \\ f_{31}(x),\ f_{32}(x),\ f_{33}(x),\ \ldots,\ f_{3n}(x),\ \ldots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_{n1}(x),\ f_{n2}(x),\ f_{n3}(x),\ \ldots,\ f_{nn}(x),\ \ldots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{array}$$

the subsequence in the nth row being convergent at each of the points $x_1, x_2, \ldots, x_n$.

Now consider what is called a "diagonal" sequence

$$f_{11}(x),\ f_{22}(x),\ \ldots,\ f_{nn}(x),\ \ldots$$

We shall prove that *this sequence uniformly converges on* $[a, b]$.

To abbreviate the notation in what follows we shall designate this (as well as the original) sequence as

$$f_1(x),\ f_2(x),\ \ldots,\ f_n(x),\ \ldots$$

(i.e. we shall use a single index instead of the double one). Take an arbitrary $\varepsilon > 0$.

Since the diagonal sequence is equicontinuous on $[a, b]$, given $\varepsilon > 0$, there is $\delta > 0$ such that whatever two points x and x_m of $[a, b]$ related by $|x - x_m| < \delta$ we may take, for all n

$$|f_n(x) - f_n(x_m)| < \frac{\varepsilon}{3}. \tag{1.35}$$

Noticing this fact we divide $[a, b]$ into a *finite* number of closed intervals of length smaller than δ. We choose a finite number n_0 of the first elements $x_1, x_2, \ldots, x_{n_0}$ of $\{x_n\}$ large enough *for each of the closed intervals to contain at least one of the points* $x_1, x_2, \ldots, x_{n_0}$.

Obviously the diagonal sequence converges at each of the points $x_1, x_2, \ldots, x_{n_0}$. Given the above $\varepsilon > 0$ therefore, we can find N such that

$$|f_{n+p}(x_m) - f_n(x_m)| < \frac{\varepsilon}{3} \tag{1.36}$$

for all $n \geqslant N$, all natural p and all $m = 1, 2, \ldots, n_0$.

Now let x be an *arbitrary* point on $[a, b]$. It would lie in one of the above closed intervals of length smaller than δ. For this x therefore there is at least one point x_m (m is one of the integers equal to $1, 2, \ldots, n_0$) satisfying the condition $|x - x_m| < \delta$.

Since the modulus of the sum of three terms is not greater than the sum of their moduli, we can write

$$|f_{n+p}(x)-f_n(x)|\leqslant|f_{n+p}(x)-f_{n+p}(x_m)|+$$
$$+|f_{n+p}(x_m)-f_n(x_m)|+|f_n(x_m)-f_n(x)|. \tag{1.37}$$

We evaluate the second term on the right-hand side of (1.37) by means of inequality (1.36) and to evaluate the first and the third term we use the fact that $|x-x_m|<\delta$ and inequality (1.35) holding for any n (and therefore for any $n+p$).

We finally have that given an arbitrary $\varepsilon>0$ there is N such that

$$|f_{n+p}(x)-f_n(x)|<\varepsilon$$

for all $n\geqslant N$, all natural p and any point x in $[a, b]$. Uniform convergence of the diagonal sequence is proved. Thus Theorem 1.12 is proved.

Remark 2. In the Arzela theorem, instead of uniform convergence of $\{f_n(x)\}$ on $[a, b]$ it is sufficient to require boundedness *at least at one point* of the interval. Indeed, the following statement is true: *if a sequence $\{f_n(x)\}$ is equicontinuous on a closed interval $[a, b]$ and bounded at least at one point x_0 of the interval, then it is uniformly bounded on $[a, b]$*. To prove this we notice that by the definition of equicontinuity, given $\varepsilon=1$ we can find $\delta>0$ such that the *oscillation* of any function $f_n(x)$ on any closed interval not greater than δ does not exceed $\varepsilon=1$. Since the entire interval $[a, b]$ can be covered with a finite number n_0 of closed intervals of length not greater than δ, the oscillation of any function $f_n(x)$ on the entire interval $[a, b]$ does not exceed the number n_0. But then the inequality $|f_n(x_0)|\leqslant A$ expressing the boundedness of $\{f_n(x)\}$ at x_0 implies the inequality $|f_n(x)|\leqslant A+n_0$ true for any point x of $[a, b]$ and expressing the uniform boundedness of the sequence under consideration on $[a, b]$.

Remark 3. We establish a sufficient test for equicontinuity: *if a sequence $\{f_n(x)\}$ consists of functions differentiable on a closed interval $[a, b]$ and if the sequence of derivatives $\{f_n'(x)\}$ is uniformly bounded on that interval, then $\{f_n(x)\}$ is equicontinuous on $[a, b]$*.

To prove this take on $[a, b]$ two arbitrary points x' and x'' and write for $f_n(x)$ on $[x', x'']$ the Lagrange formula (see Section 8.9 of [1]).

By the Lagrange theorem, there is a point ξ_n on $[x', x'']$ such that

$$|f_n(x')-f_n(x'')|=|f_n'(\xi_n)|\cdot|x'-x''|. \tag{1.38}$$

Since $\{f_n'(x)\}$ is uniformly bounded on $[a, b]$, there is a constant A such that for all n

$$|f_n'(\xi_n)|\leqslant A. \tag{1.39}$$

Substituting (1.39) into (1.38) we get

$$|f_n(x')-f_n(x'')|\leqslant A\,|x'-x''|. \tag{1.40}$$

Choose any $\varepsilon > 0$. Then, taking $\delta = \varepsilon/A$ and using (1.40), we get, for all n and for all x' and x'' of $[a, b]$ related by $|x' - x''| < \delta$, the inequality

$$|f_n(x') - f_n(x'')| < \varepsilon.$$

The equicontinuity of $\{f_n(x)\}$ is thus proved.

As an example consider a sequence $\left\{\frac{\sin nx}{n}\right\}$. This sequence is equicontinuous on any closed interval $[a, b]$, for the sequence of derivatives $\{\cos nx\}$ is uniformly bounded on any closed interval $[a, b]$.

Remark 4. The concept of equicontinuity can be formulated not only with respect to the closed interval $[a, b]$, but also with respect to the open interval, half-open interval, half-line, an infinite straight line and with respect to any set dense in itself in general*. Moreover, this notion may be applied not to the sequence of functions but to any infinite number of functions.

1.4. POWER SERIES

1.4.1. A power series and its domain of convergence. A *power series* is a functional series of the form

$$a_0 + \sum_{k=1}^{\infty} a_k x^k = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots, \tag{1.41}$$

where $a_0, a_1, a_2, \ldots, a_n, \ldots$ are constant real numbers called the *coefficients* of the series (1.41). Let us try and see what the *domain of convergence* of any power series looks like.

Notice that *any power series converges at a point $x = 0$*, there existing pover series converging only at that point (for example, the series $\sum_{k=1}^{\infty} k! \cdot x^k$ does).

Form the following number sequence using the coefficients a_n of (1.41):

$$\{\sqrt[n]{|a_n|}\} \quad (n = 1, 2, \ldots). \tag{1.42}$$

There may occur two cases: (1) the sequence (1.42) is *unbounded*; (2) (1.42) is *bounded*.

In case (2) the sequence (1.42) has a *finite upper limit* (see Section 3.4.3 in [1]) which we denote by L. It should be stressed that the upper limit L is automatically *nonnegative* (for all the elements of (1.42) and therefore any limit point of this sequence are nonnegative).

---

* The Arzela theorem remains valid if we replace the closed interval $[a, b]$ by any closed bounded set in the theorem.

Summarizing we come to the conclusion that the following three cases are possible: (I) the sequence (1.42) is unbounded; (II) (1.42) is bounded and has a finite upper limit $L > 0$; (III) (1.42) is bounded and has an upper limit $L = 0$.

Now we prove the following remarkable statement.

Theorem 1.13 (Cauchy-Hadamard).

I. *If the sequence* (1.42) *is not bounded, then the power series* (1.41) *converges only when* $x = 0$.

II. *If the sequence* (1.42) *is bounded and has an upper limit* $L > 0$, *then the series* (1.41) *absolutely converges for values of* x *satisfying the inequality* $|x| < 1/L$ *and diverges for values of* x *satisfying* $|x| > 1/L$.

III. *If the sequence* (1.42) *is bounded and its upper limit* $L = 0$, *then the series* (1.41) *absolutely converges for all values of* x.

Proof.

I. Let (1.42) be unbounded. Then for $x \neq 0$ so is the sequence

$$|x| \cdot \sqrt[n]{|a_n|} = \sqrt[n]{|a_n \cdot x^n|},$$

i.e. this sequence has elements *with arbitrarily large n satisfying*

$$\sqrt[n]{|a_n \cdot x^n|} > 1 \quad \text{or} \quad |a_n \cdot x^n| > 1.$$

But this means that the necessary condition of convergence fails for the series (1.41) (with $x \neq 0$) (see Section 13.1.2 in [1]), i.e. (1.41) diverges when $x \neq 0$.

II. Let the sequence (1.42) be bounded and let its upper limit $L > 0$. We prove that (1.41) absolutely converges when $|x| < 1/L$ and diverges when $|x| > 1/L$.

(a) First choose any x satisfying $|x| < 1/L$. Then there is $\varepsilon > 0$ such that $|x| < 1/(L + \varepsilon)$. By the properties of the upper limit all elements $\sqrt[n]{|a_n|}$ beginning with some n satisfy

$$\sqrt[n]{|a_n|} < L + \frac{\varepsilon}{2}.$$

Thus, beginning with this n,

$$\sqrt[n]{|a_n \cdot x^n|} = |x| \cdot \sqrt[n]{|a_n|} < \frac{L + \frac{\varepsilon}{2}}{L + \varepsilon} < 1,$$

i.e. the series (1.41) absolutely converges by the Cauchy test (see Section 13.2.3 in [1]).

(b) Now choose any x satisfying $|x| > 1/L$.

Then there is $\varepsilon > 0$ such that $|x| > 1/(L - \varepsilon)$. By the definition of the upper limit we may choose a subsequence $\left\{\sqrt[n_k]{|a_{n_k}|}\right\}$ $(k = 1, 2, \ldots)$ of the sequence (1.42) converging to L.

But this means that, beginning with some k,

$$L-\varepsilon<\sqrt[n_k]{|a_{n_k}|}<L+\varepsilon.$$

Thus, beginning with that k,

$$\sqrt[n_k]{|a_{n_k}\cdot x^{n_k}|}=|x|\cdot\sqrt[n_k]{|a_{n_k}|}>\frac{L-\varepsilon}{L-\varepsilon}=1$$

or

$$|a_{n_k}\cdot x^{n_k}|>1,$$

i.e. the necessary condition for the convergence of the series (1.41) fails and the series diverges.

III. Let the sequence (1.42) be bounded and let its upper limit $L=0$. We prove that (1.41) absolutely converges for any x.

We choose an arbitrary $x\neq 0$ (when $x=0$ the series (1.41) automatically converges absolutely). Since the upper limit $L=0$ and the sequence (1.42) cannot have negative limit points, the number $L=0$ is the *only* limit point and therefore it is the limit of that sequence, i.e. (1.42) is infinitesimal.

But then for any positive number $1/2\ |x|$ there is an integer beginning with which

$$\sqrt[n]{|a_n|}<\frac{1}{2|x|}.$$

Therefore, beginning with that integer,

$$\sqrt[n]{|a_n x^n|}=|x|\cdot\sqrt[n]{|a_n|}<\frac{1}{2}<1,$$

i.e. (1.41) absolutely converges by the Cauchy test (see Section 13.2.3 in [1]). This completes the proof of the theorem.

The theorem leads to the following fundamental statement.

Theorem 1.14. For every power series (1.41), *unless it is one converging only at a point $x=0$, there is a positive number R (possibly equal to infinity) such that that series absolutely converges when $|x|<R$ and diverges when $|x|>R$.*

The number R is called the *radius of convergence* of the series and the interval $(-R, R)$ is the *interval of convergence* of the series. The radius of convergence is computed from the formula

$$R=\frac{1}{\overline{\lim\limits_{n\to\infty}}\sqrt[n]{|a_n|}} \tag{1.43}$$

(when $\overline{\lim\limits_{n\to\infty}}\sqrt[n]{|a_n|}=0,\ R=\infty$).

Remark 1. At the end points of the interval of convergence, i.e. at the points $x = -R$ and $x = R$, a power series may be both convergent and divergent*.

Thus the series $1 + \sum_{k=1}^{\infty} x^k$ has a radius of convergence equal to unity and the interval of convergence of the form $(-1, 1)$ and converges at the end points of the interval.

The series $\sum_{k=1}^{\infty} \frac{x^k}{k^2}$ has the same interval of convergence, $(-1, 1)$, but converges at both end points of the interval.

Remark 2. All the results of this subsection hold for a series (1.41) in which the real variable x is replaced by a complex variable z.

For such a series we can establish the existence of a positive number R such that the series absolutely converges when $|z| < R$ and diverges when $|z| > R$.

To compute R we use formula (1.43). The number R is called the *radius of convergence* of the series and the region $|z| < R$ is the *circle of convergence* of the series.

1.4.2. Continuity of the sum of a power series. Let a power series (1.41) have the radius of convergence $R > 0$.

Lemma 2. Given any positive number r satisfying the condition $r < R$, the series (1.41) uniformly converges on a closed interval $[-r, r]$, i.e. when $|x| \leqslant r$.

Proof. By Theorem 1.14 the series (1.41) absolutely converges when $x = r$, i.e.

$$|a_0| + \sum_{k=1}^{\infty} |a_k| \cdot r^k$$

converges. But this number series serves as a majorant of (1.41) for all x of $[-r, r]$. By the Weierstrass test, (1.41) converges uniformly on $[-r, r]$. Thus the lemma is proved.

Corollary. Under the hypotheses of Lemma 2 the sum of a series (1.41) *is a function continuous on a closed interval $[-r, r]$* (by Theorem 1.7).

Theorem 1.15. The sum of a power series is a continuous function in the interval of convergence of the series.

Proof. Let $S(x)$ be the sum of a power series (1.41) and let R be its radius of convergence. We take any x in the interval of convergence, i.e. such that $|x| < R$. There is always a number r such that $|x| < r < R$. By the corollary of Lemma 2, $S(x)$ is con-

* Note the following theorem of Abel: *if the power series (1.41) converges when $x = R$, then its sum $S(x)$ is continuous at the point R from the left.* We may assume without loss of generality that $R = 1$, but in this form Abel's theorem (asserting in fact the regularity of the Abel-Poisson summation method) is proved in Supplement 3 to Chapter 13 in [1].

tinuous on $[-r, r]$. Therefore $S(x)$ is continuous at x too. The theorem is proved.

1.4.3. Term-by-term integration and term-by-term differentiation of a power series.

Theorem 1.16. If $R>0$ is the radius of convergence of a power series (1.41) and x satisfies the condition $|x|<R$, then (1.41) can be integrated term by term over the closed interval $[0, x]$. The resulting series has the same radius of convergence R that the original series has.

Proof. For any x satisfying the condition $|x|<R$ we can find r such that $|x|<r<R$. By Lemma 2, (1.41) converges uniformly on $[-r, r]$ and hence on $[0, x]$. But then, by Theorem 1.8, that series can be integrated over $[0, x]$ term by term.

Integrating term by term yields a power series

$$a_0x+\frac{a_1}{2}x^2+\ldots+\frac{a_{n-1}}{n}x^n+\ldots$$

whose radius of convergence, by Theorem 1.14, is the inverse of the upper limit of the sequence

$$\sqrt[n]{\frac{|a_{n-1}|}{n}}=\frac{\sqrt[n]{|a_{n-1}|}}{\sqrt[n]{n}}. \tag{1.44}$$

Since the upper limit of (1.44) is the same that (1.42) has*, the theorem is proved.

Theorem 1.17. A power series (1.41) can be differentiated term by term in its interval of convergence. The resulting series has the same radius of convergence R that the original series has.

Proof. It is sufficient (by Theorem 1.9 and Lemma 2) to prove only the second statement of the theorem.

Differentiating (1.41) term by term yields a series

$$a_1+2\cdot a_2\cdot x+\ldots+n\cdot a_n\cdot x^{n-1}+(n+1)\cdot a_{n+1}\cdot x^n+\ldots$$

whose radius of convergence (by Theorem 1.14) is the inverse of the upper limit of the sequence

$$\{\sqrt[n]{(n+1)\,|a_{n+1}|}\}. \tag{1.45}$$

---

* For $\lim\limits_{n\to\infty}\sqrt[n]{n}=1$, $\varlimsup\limits_{n\to\infty}\sqrt[n]{|a_{n-1}|}=\varlimsup\limits_{n\to\infty}\sqrt[n+1]{|a_n|}=$

$$=\varlimsup_{n\to\infty}\left[\sqrt[n]{|a_n|}\right]^{\frac{n}{n+1}}=\varlimsup_{n\to\infty}\left[\sqrt[n]{|a_n|}\right].$$

Since (1.45) has the same upper limit that (1.42) has*, the theorem is proved.

Corollary. A power series can be differentiated term by term as many times as we like in its interval of convergence.

The series resulting after n-fold term-by-term differentiation of the original series has the same radius of convergence that the original series has.

1.5. POWER SERIES EXPANSION OF FUNCTIONS

1.5.1. Power series expansion of a function.

Definition 1. We shall say that a function $f(x)$ can be expanded into a power series on an open interval $(-R, R)$ (on a set $\{x\}$) if there is a power series converging to $f(x)$ on $(-R, R)$ (on $\{x\}$).

The following statements are true.

1°. *For a function $f(x)$ to allow expansion into a power series on an open interval $(-R, R)$, it is necessary that it should have continuous derivatives of any order on $(-R, R)$***.

Indeed, a power series can be differentiated term by term as many times as we please in its interval of convergence which in any case contains the open interval $(-R, R)$, all the resulting series converging in the same interval of convergence (Theorem 1.17).

But then the sums of the series obtained by differentiating as many times as we please (by Theorem 1.15) are functions continuous in the interval of convergence and are therefore continuous on $(-R, R)$.

2°. *If a function $f(x)$ can be expanded into a power series on an open interval $(-R, R)$, this can only be done in a unique way.*

Indeed, let $f(x)$ be a function that can be expanded on $(-R, R)$ into a power series (1.41).

Differentiating (1.41) term by term n times (which can trivially be done in the interval $(-R, R)$) we get

$$f^{(n)}(x) = a_n \cdot n! + a_{n+1} \cdot (n+1)\, x! + \ldots .$$

---

* For $\lim\limits_{n\to\infty} \sqrt[n]{n-1} = 1$, $\overline{\lim\limits_{n\to\infty}} \sqrt[n]{|a_{n+1}|} =$

$$= \overline{\lim_{n\to\infty}} \sqrt[n-1]{|a_n|} = \overline{\lim_{n\to\infty}} \left[\sqrt[n]{|a_n|}\right]^{\frac{n}{n-1}} = \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|}.$$

** Note that there are functions that have continuous derivatives of any order on $(-R, R)$ but cannot be expanded into a power series on that interval. As an illustration

$$f(x) = \begin{cases} e^{-1/x^2} & \text{when } x \neq 0, \\ 0 & \text{when } x = 0. \end{cases}$$

From which for $x = 0$ we find

$$f^{(n)}(0) = a_n \cdot n!$$

or

$$a_n = \frac{f^n(0)}{n!}. \tag{1.46}$$

Thus the coefficients of (1.41) into which we can expand $f(x)$ are uniquely defined by formula (1.46).

Suppose now that $f(x)$ has continuous derivatives of any order on $(-R, R)$.

Definition 2. A power series (1.41) *whose coefficients are defined by formula* (1.46) *are called a* **Taylor series** *of the function $f(x)$.*

Statement 2° leads us to the following statement.

3°. *If a function $f(x)$ can be expanded into a power series on an open interval $(-R, R)$, that series is a Taylor series of the function $f(x)$.*

In conclusion we formulate the following statement immediate from Section 8.14 of [1].

4°. *For a function $f(x)$ to allow expansion into a Taylor series on an open interval $(-R, R)$ (on a set $\{x\}$) it is necessary and sufficient that the remainder in the Maclaurin formula for that function should tend to zero on $(-R, R)$ (on $\{x\}$).*

1.5.2. Taylor expansion of some elementary functions. It was proved in [1] (see Section 8.15.2) that the remainders in the Maclaurin formula for the functions e^x, $\cos x$, and $\sin x$ tend to zero on the entire infinite straight line and that the remainder in the Maclaurin formula for the function $\ln(1+x)$ tends to zero on the half-open interval $-1 < x \leqslant +1$.

By Statement 4° of the preceding subsection this brings us to the following expansions:

$$e^x = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!},$$

$$\cos x = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!},$$

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!},$$

$$\ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^n}{n}.$$

The first three of these converge for all values of x, while the last one does for the values of x in $-1 < x \leqslant 1$.

Now we discuss power series expansion of the function $(1+x)^\alpha$ or what is called a *binomial series.*

If $f(x) = (1+x)^\alpha$, then

$f^{(n)}(x) = \alpha(\alpha-1)(\alpha-2)\ldots(\alpha-n+1)\cdot(1+x)^{\alpha-n}$.

The Maclaurin formula with remainder in the Cauchy form is therefore of the form (see Section 8.14 of [1])

$$(1+x)^\alpha = 1 + \sum_{k=1}^{n} \frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!} x^k + R_{n+1}(x), \qquad (1.47)$$

where

$$R_{n+1}(x) = \frac{(1-\theta)^n}{n!} x^{n+1}\cdot f^{(n+1)}(\theta x) =$$

$$= \frac{(1-\theta)^n}{n!}\cdot x^{n+1}\cdot\alpha(\alpha-1)\ldots(\alpha-n)(1+\theta x)^{\alpha-n-1} =$$

$$= \left(\frac{1-\theta}{1-\theta x}\right)^n\cdot\frac{(\alpha-1)(\alpha-2)\ldots(\alpha-n)}{n!}\cdot\alpha(1+\theta x)^{\alpha-1}\cdot x^{n+1} \qquad (1.48)$$

(θ is some number in the interval $0<\theta<1$).

We first show that when $\alpha > 0$ the remainder $R_{n+1}(x)$ tends to zero (as $n\to\infty$) everywhere on $-1<x<1$.

Indeed, the elements of $\left\{\left(\frac{1-\theta}{1+\theta x}\right)^n\right\}$ are all not greater than unity anywhere on $-1<x<1$; $\left\{\frac{(\alpha-1)(\alpha-2)\ldots(\alpha-n)}{n!}\right\}$ is bounded for any given $\alpha>0$*; the number $\alpha(1+\theta x)^{\alpha-1}$ is defined for any given $\alpha>0$ and any x in the interval $-1<x<+1$; finally the sequence $\{x^{n+1}\}$ is infinitesimal for any x in $-1<$ $<x<1$.

Thus by (1.48) the remainder $R_{n+1}(x)$ tends to zero for any given $\alpha>0$ and any x in $-1<x<1$.

By (1.47) therefore, *when $\alpha>0$, everywhere on $-1<x<1$ we have*

$$(1+x)^\alpha = 1 + \sum_{k=1}^{\infty} \frac{\alpha(\alpha-1)\ldots(\alpha-k+1)}{k!} x^k. \qquad (1.49)$$

Now we prove that *when $\alpha>0$ the series on the right of (1.49) uniformly converges to $(1+x)^\alpha$ on the closed interval $-1\leqslant x\leqslant 1$.*

Everywhere on this interval the series is majorized by the following number series:

$$\sum_{k=1}^{\infty} \frac{|\alpha|\cdot|1-\alpha|\ldots|k-1-\alpha|}{k!}. \qquad (1.50)$$

---

* All the elements of the sequence are bounded in absolute value by the number $\frac{(\alpha-1)(\alpha-2)\ldots(\alpha-[\alpha])}{[\alpha]!}$, where $[\alpha]$ is the integral part of α.

By the Weierstrass test, to establish uniform convergence on $-1 \leqslant x \leqslant 1$ of the series on the right of (1.49) it is sufficient to prove the convergence of the majorant series (1.50).

Denote the kth term of (1.50) by p_k. Then for all large enough k we get

$$\frac{p_{k+1}}{p_k} = \frac{k-\alpha}{k+1} = 1 - \frac{1+\alpha}{k+1}. \tag{1.51}$$

This yields

$$\lim_{k\to\infty} k\left(1 - \frac{p_{k+1}}{p_k}\right) = (1+\alpha) \cdot \lim_{k\to\infty} \frac{k}{k+1} = 1+\alpha > 1,$$

i.e. (1.50) converges by the *Raabe test* (see Section 13.2.5 of [1]).

We have thereby proved that when $\alpha > 0$ the series on the right of (1.49) converges uniformly on $-1 \leqslant x \leqslant 1$. It remains to prove that that series converges on $-1 \leqslant x \leqslant 1$ to $(1+x)^\alpha$.

By what was proved above the sum of the series on the right of (1.49), $S(x)$, and the function $(1+x)^\alpha$ *coincide everywhere on* $-1 < x < 1$. In addition both functions, $S(x)$ and $(1+x)^\alpha$, *are continuous on* $-1 \leqslant x \leqslant 1$ (the function $S(x)$ as the sum of a uniformly convergent series of continuous functions; the continuity of $(1+x)^\alpha$ is obvious when $\alpha > 0$).

But then the values of $S(x)$ and $(1+x)^\alpha$ at $x = -1$ and $x = 1$ must coincide, i.e. the series on the right of (1.49) uniformly converges to $(1+x)^\alpha$ on $-1 \leqslant x \leqslant 1$.

1.5.3. Some elementary facts about the functions of a complex variable. As already noted above, we may extend to the case of a power series in a complex variable z

$$a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n + \ldots$$

Theorems 1.13 and 1.14 (on the existence and value of the radius of convergence). Series of this type are used to define the functions of a complex variable z.

The functions e^z, $\cos z$, and $\sin z$ of z are defined to be the sums of the following series:

$$e^z = 1 + \sum_{n=1}^{\infty} \frac{z^n}{n!}, \tag{1.52}$$

$$\cos z = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n \cdot z^{2n}}{(2n)!}, \tag{1.53}$$

$$\sin z = \sum_{n=0}^{\infty} \frac{(-1)^n \cdot z^{2n+1}}{(2n+1)!} \tag{1.54}$$

It is easy to verify that these three series absolutely converge for all values of z (their radius of convergence $R = \infty$).

Now we establish the relation between the functions e^z, $\cos z$, and $\sin z$.

Replacing z by iz in formula (1.52) we get

$$e^{iz}=1+iz+\frac{(iz)^2}{2!}+\frac{(iz)^3}{3!}+\frac{(iz)^4}{4!}+\frac{(iz)^5}{5!}+\ldots=$$

$$=\left(1-\frac{z^2}{2!}+\frac{z^4}{4!}-\ldots\right)+i\left(z-\frac{z^3}{3!}+\frac{z^5}{5!}-\ldots\right). \qquad (1.55)$$

Comparing the right-hand side of (1.55) with the expansions (1.53) and (1.54) we arrive at the following remarkable formula:

$$e^{iz}=\cos z+i\cdot\sin z. \qquad (1.56)$$

Formula (1.56) plays a fundamental part in the theory of functions of a complex variable and is called *Euler's formula*.

Setting in Euler's formula the variable z first equal to a real x and then to a real $-x$ we obtain the following two formulas:

$$e^{ix}=\cos x+i\cdot\sin x,\quad e^{-ix}=\cos x-i\sin x.$$

Adding and subtracting these we obtain formulas expressing $\cos x$ and $\sin x$ in terms of an exponential function:

$$\begin{cases}\cos x=\dfrac{e^{ix}+e^{-ix}}{2},\\[2mm] \sin x=\dfrac{e^{ix}-e^{-ix}}{2i}.\end{cases} \qquad (1.57)$$

In conclusion we shall discuss the definition of the logarithmic function $w=\ln z$ of the complex variable z. It is natural to define this function as the inverse of the exponential, i.e. by the relation $z=e^w$. Putting $w=u+iv$, $z=x+iy$ we intend to express u and v in terms of $z=x+iy$.

From the relation

$$z=x+iy=e^{u+iv}=e^u(\cos v+i\sin v)$$

we get, using the concepts of the modulus and argument of a complex number (see formula (7.6) in [1]),

$$|z|=\sqrt{x^2+y^2}=e^u,\quad \arg z=v-2\pi k,$$

where

$$k=0,\ \pm1,\ \pm2,\ \ldots$$

From these equations we find that

$$u=\ln|z|=\ln\sqrt{x^2+y^2},$$

$$v=\arg z+2\pi k\quad (k=0,\ \pm1,\ \pm2,\ldots)$$

or, finally, that

$$\ln z=\ln|z|+i(\arg z+2\pi k),\ \text{where } k=0,\ \pm1,\ \pm2,\ \ldots. \qquad (1.58)$$

Formula (1.58) shows that the logarithmic function *is not single-valued* in the complex domain: its imaginary part has an infinite number of values corresponding to different $k=0,\ \pm1,\ \pm2,\ \ldots$ for the same value of z.

It is easy to realize that a similar situation will arise when inverse trigonometric functions are defined in the complex domain.

1.5.4. Uniform approximation of a continuous function by polynomials (the Weierstrass theorem). In this subsection we shall prove the fundamental theorem due to Weierstrass who established it in 1895.

Theorem 1.18 **(*Weierstrass*).** *If a function $f(x)$ is continuous on a closed interval $[a, b]$, then there is a polynomial sequence $\{P_n(x)\}$ uniformly converging on $[a, b]$ to $f(x)$, i.e. given any $\varepsilon > 0$, there is a polynomial $P_n(x)$ with an integer n dependent on ε such that*

$$|P_n(x) - f(x)| < \varepsilon$$

at once for all x in $[a, b]$.

Restated, *a function $f(x)$ continuous on a closed interval $[a, b]$ can be uniformly approximated on $[a, b]$ by a polynomial with a preassigned accuracy ε.*

Proof. Without loss of generality we may consider a closed interval $[0, 1]$ instead of $[a, b]$*. Moreover, it is sufficient to prove the theorem for a continuous function $f(x)$ vanishing at the ends of $[0, 1]$, i.e. satisfying the conditions $f(0) = 0$ and $f(1) = 0$. Indeed, if $f(x)$ did not satisfy these conditions, then on putting

$$g(x) = f(x) - f(0) - x[f(1) - f(0)]$$

we should obtain a function $g(x)$ continuous on $[0, 1]$ and satisfying the conditions $g(0) = 0$ and $g(1) = 0$, and it would follow from the possibility of representing $g(x)$ as the limit of a uniformly convergent sequence of polynomials that $f(x)$ can also be represented as the limit of a uniformly convergent sequence of polynomials (for the difference $f(x) - g(x)$ is a first-degree polynomial).

So let $f(x)$ be a function continuous on $[0, 1]$ and satisfying the conditions $f(0) = 0$, $f(1) = 0$. We may extend such a function to the entire infinite straight line by setting it equal to zero outside $[0, 1]$ and state that a function extended in this way *is uniformly continuous on the entire infinite straight line.*

Consider the following particular sequence of non-negative polynomials of degree $2n$:

$$Q_n(x) = c_n(1 - x^2)^n \quad (n = 1, 2, \ldots), \tag{1.59}$$

each having a constant c_n chosen so that

$$\int_{-1}^{1} Q_n(x)\,dx = 1 \quad (n = 1, 2, \ldots). \tag{1.60}$$

We majorize c_n without computing its exact value.

* Since one of the intervals transforms to the other by linear substitution $x = (b - a)t + a$.

To do this we notice that for any $n = 1, 2, \ldots$ and all x in $[0, 1]$ we have the inequality*

$$(1 - x^2)^n \geqslant 1 - nx^2. \tag{1.61}$$

Using (1.61) and the fact that $1/\sqrt{n} \leqslant 1$ for any $n \geqslant 1$ we have

$$\int_{-1}^{1} (1-x^2)^n\,dx = 2\int_0^1 (1-x^2)^n\,dx \geqslant 2\int_0^{1/\sqrt{n}} (1-x^2)^n\,dx \geqslant$$

$$\geqslant 2\int_0^{1/\sqrt{n}} (1-nx^2)\,dx = \frac{4}{3}\,\frac{1}{\sqrt{n}} > \frac{1}{\sqrt{n}}. \tag{1.62}$$

From (1.59), (1.60), and (1.62) we derive that given any $n = 1, 2, \ldots$ the following upper estimate holds for c_n:

$$c_n < \sqrt{n}. \tag{1.63}$$

From (1.63) and (1.59) it follows that given any $\delta > 0$, for all x in $\delta \leqslant x \leqslant 1$ we have

$$0 \leqslant Q_n(x) \leqslant \sqrt{n}\,(1-\delta^2)^n \tag{1.64}$$

From (1.64) we deduce tha *given any fixed $\delta > 0$ the sequence of nonnegative polynomials $\{Q_n(x)\}$ converges to zero uniformly on* $\delta \leqslant x \leqslant 1$**.

Now for any x in $0 \leqslant x \leqslant 1$ we put

$$P_n(x) = \int_{-1}^{1} f(x+t)\,Q_n(t)\,dt \tag{1.65}$$

and show that given any $n = 1, 2, \ldots$ the function $P_n(x)$ is a polynomial of degree $2n$, with $\{P_n(x)\}$ the desired sequence of polynomials uniformly converging on $0 \leqslant x \leqslant 1$ to $f(x)$.

---

* This inequality follows from the fact that for any $n \geqslant 1$ the function $\varphi(x) = (1 - x^2)^n - (1 - nx^2)$ *is nonnegative everywhere on* $0 \leqslant x \leqslant 1$, for it vanishes when $x = 0$ and has a nonnegative derivative $\varphi'(x) = 2nx\,[1 - (1 - x^2)^{n-1}]$ everywhere on $0 \leqslant x \leqslant 1$.

** Indeed, it suffices to prove that the sequence $a_n = (1 - \delta^2)^n \cdot \sqrt{n}$ converges to zero, and this follows, for example, from the fact that since

$$\lim_{n\to\infty} \sqrt[n]{a_n} = (1-\delta^2) \lim_{n\to\infty} n^{1/2n} = (1-\delta^2) < 1,$$

the series $\sum_{n=1}^{\infty} a_n$ converges by the Cauchy test (see Theorem 13.6 in [1]).

Since $f(x)$, the function under study, is zero outside $[0, 1]$, for any x of $[0, 1]$ the integral (1.65) can be written as

$$P_n(x) = \int_{-x}^{1-x} f(x+t)\, Q_n(t)\, dt.$$

Replacing t by $t - x$ in the last integral we get

$$P_n(x) = \int_{0}^{1} f(t)\, Q_n(t-x)\, dt. \tag{1.66}$$

It is clear from (1.66) and (1.59) that $P_n(x)$ is a polynomial of degree $2n$.

It remains to prove that $\{P_n(x)\}$ converges to $f(x)$ uniformly on $0 \leqslant x \leqslant 1$.

We take an arbitrary $\varepsilon > 0$. For given ε, by uniform continuity of $f(x)$ on the entire infinite straight line, we can find $\delta > 0$ such that

$$| f(x) - f(y) | < \frac{\varepsilon}{2} \text{ when } | x - y | < \delta. \tag{1.67}$$

We further notice that since $f(x)$ is continuous on $[0, 1]$, it is also bounded on that interval and therefore everywhere on the infinite straight line. This means that there is a constant A such that for all x

$$| f(x) | \leqslant A. \tag{1.68}$$

Using (1.60), (1.64), (1.67), and (1.68) and the non-negativeness of $Q(x)$ we evaluate the difference $P_n(x) - f(x)$.

For all x in $0 \leqslant x \leqslant 1$ we have

$$| P_n(x) - f(x) | = \left| \int_{-1}^{1} [f(x+t) - f(x)]\, Q_n(t)\, dt \right| \leqslant$$

$$\leqslant \int_{-1}^{1} | f(x+t) - f(x) |\, Q_n(t)\, dt \leqslant 2A \int_{-1}^{-\delta} Q_n(t)\, dt +$$

$$+ \frac{\varepsilon}{2} \int_{-\delta}^{\delta} Q_n(t)\, dt + 2A \int_{\delta}^{1} Q_n(t)\, dt \leqslant$$

$$\leqslant 4A \sqrt{n} \cdot (1-\delta^2)^n + \frac{\varepsilon}{2}.$$

To complete the proof of the theorem it suffices to notice that for all large enough n

$$4A\sqrt{n}\,(1-\delta^2)^n < \frac{\varepsilon}{2}.$$

Corollary. If not only the function $f(x)$ itself, but also its derivatives up to some order k inclusively are continuous on a closed interval $[0, 1]$*, *then there is a sequence of polynomials* $\{P_n(x)\}$ *such that each of the sequences* $\{P_n(x)\}$, $\{P'_n(x)\}$, $\ldots$, $\{P_n^{(k)}(x)\}$ *converges uniformly on* $[0, 1]$ *to* $f(x)$, $f'(x)$, $\ldots$, $f^{(k)}(x)$ respectively.

Indeed, we may assume without loss of generality that each of the functions $f(x)$, $f'(x)$, $\ldots$, $f^{(k)}(x)$ vanishes when $x = 0$ and when $x = 1$**, and under such conditions $f(x)$ can be extended to the entire infinite straight line by setting it equal to zero outside $[0, 1]$, so that the extended function and all of its derivatives up to order k inclusively turn out to be *uniformly continuous on the entire infinite straight line.*

But then, denoting by $P_n(x)$ the same polynomial (1.65) as above and repeating the arguments used in proving Theorem 1.18, we establish that each of the differences

$$P_n(x)-f(x),\ P'_n(x)-f'(x),\ \ldots,\ P_n^{(k)}(x)-f^{(k)}(x)$$

is an infinitesimal uniform in x on $0 \leqslant x \leqslant 1$.

Remark 1. The above proof can be easily generalized to the case of a function of m variables $f(x_1, x_2, \ldots, x_m)$ continuous in an m-dimensional cube $0 \leqslant x_i \leqslant 1$ $(i = 1, 2, \ldots, m)$.

In full analogy with Theorem 1.18 it can be proved that for such a function $f(x_1, x_2, \ldots, x_m)$ there is a sequence of polynomials in m variables $x_1, x_2, \ldots, x_m$ uniformly converging to it in the m-dimensional cube.

Remark 2. Notice that the polynomials of Theorem 1.18 may be replaced by functions of a more general nature while retaining the statement about the possibility of uniform approximation by such functions of any continuous function f.

Let us agree to say that an arbitrary collection A of functions defined on some set E is *an algebra* if*** (1) $f+g \in A$; (2) $f\cdot g \in A$; (3) $\alpha\cdot f \in A$ for arbitrary $f \in A$ and $g \in A$ and for any real α.

In other words, an algebra is a collection of functions closed under addition and multiplication of functions and under multiplication of functions by real numbers.

---

* Of course we may take $[a, b]$ instead of $[0,1]$.

** If $f(x)$ failed to satisfy these conditions, we should find a polynomial $\overline{P_k}(x)$ of degree $2k$ such that these conditions would hold for $g(x) = f(x) - \overline{P_k}(x)$.

*** We recall that the symbol $f \in A$ means 'f belongs to A'.

If for each point x of E there is some function $g \in A$ such that $g(x) \neq 0$, the algebra A is said *to vanish at none of the points x of E.*

A collection A of functions defined of a set E is said *to separate the points of E* if for any two distinct points x_1 and x_2 of the set we can find a function of A such that $f(x_1) \neq f(x_2)$.

We come to the following remarkable statement called the *Stone-Weierstrass theorem**.

*Let A be an algebra of functions continuous on a compact** set E that separates the points of E and vanishes at no point of that set. Then every function $f(x)$ continuous on E can be represented as the limit of a uniformly convergent sequence of functions of A.*

---

* M. Stone is a modern American mathematician.

** Recall that a compact set is a closed bounded set

CHAPTER 2

DOUBLE INTEGRALS AND n-FOLD MULTIPLE INTEGRALS

In Volume 1 we discussed physical and geometrical problems leading to the concept of single definite integral.

Typical problems of this kind are the problem of calculating the mass of an inhomogeneous rod from the known linear density of the rod and the problem of calculating the area of a curvilinear trapezoid (i.e. the area under the graph of a nonnegative function $y = f(x)$ on a closed interval $[a, b]$).

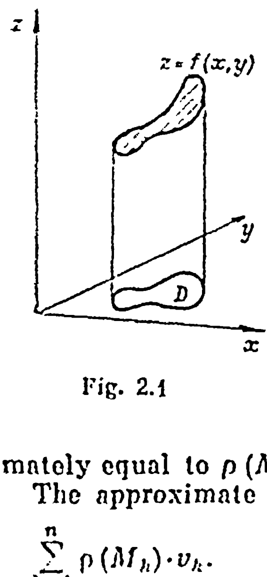


Fig. 2.1

It is easy to cite similar "multidimensional" problems leading to the notion of double or triple integral.

Thus the problem of calculating the mass of an inhomogeneous body T from the known volume density $\rho(M)$ of the body leads us in a natural way to the concept of triple integral.

To compute the mass of the body T, divide T into sufficiently small parts $T_1, T_2, \ldots, T_n$. We may roughly assume the volume density $\rho(M)$ of each part T_k to be constant and equal to $\rho(M_k)$, where M_k is some point of T_k. In this case the mass of each T_k will be approximately equal to $\rho(M_k)\cdot v_k$, where v_k is the volume of T_k.

The approximate value of the mass of the entire body T is

$$\sum_{k=1}^{n} \rho(M_k)\cdot v_k.$$

It is natural to define the exact value of the mass as the limit of the above sum as every part T_k decreases indefinitely.* We may take this limit as the definition of triple integral of $\rho(M)$ over a three-dimensional region T.

In a quite similar way we may consider the geometrical problem of calculating the volume of the so-called curved-base cylinder (i.e. the volume of the body (Fig. 2.1) lying under the graph of the nonnegative function $z = f(x, y)$ in some two-dimensional domain D).

* Of course the words "decreases indefinitely" should be made more precise.

This problem leads us to the notion of double integral of a function $f(x, y)$ over a two-dimensional domain D.

In this chapter we present the theory of double, triple and n-fold multiple integrals in general.

To make more effective use of the analogy of the single integral we first introduce the concept of double integral for the rectangle and only then do we introduce the double integral over an arbitrary domain using both rectilinear and absolutely arbitrary *subdivision* of the domain.

2.1. DEFINITION AND EXISTENCE OF A DOUBLE INTEGRAL

2.1.1. Definition of a double integral for the rectangle. Let $f(x, y)$ be a function defined everywhere on a rectangle $R = [a \leqslant x \leqslant b] \times [c \leqslant y \leqslant d]$ (Fig. 2.2).

We divide the closed interval $a \leqslant x \leqslant b$ into n subdivisions using the points $a = x_0 < x_1 < x_2 < \ldots < x_n = b$ and $c \leqslant y \leqslant d$ into p subdivisions using the points $c = y_0 < y_1 < y_2 < \ldots < y_p = d$.

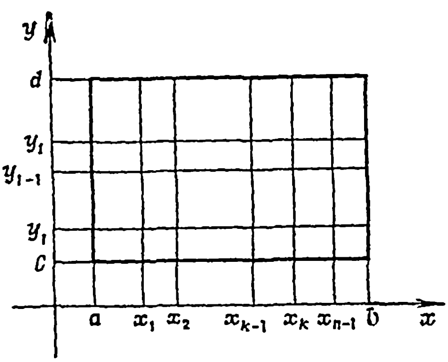


Fig. 2.2

Corresponding to subdivision by straight lines parallel to the x and y axes (see Fig. 2.2) is a subdivision of a rectangle R into $n \cdot p$ subrectangles

$$R_{kl} = [x_{k-1} \leqslant x \leqslant x_k] \times [y_{l-1} \leqslant y \leqslant y_l]$$
$$(k = 1, 2, \ldots, n; \; l = 1, 2, \ldots, p).$$

This subdivision of R will be designated by the symbol T.

Throughout what follows in this chapter, by "rectangle" we mean a rectangle with sides parallel to the coordinate axes.

On every subrectangle R_{kl} we choose a point (ξ_k, η_l). Putting $\Delta x_k = x_k - x_{k-1}$, $\Delta y_l = y_l - y_{l-1}$, denote by ΔR_{kl} the area of a rectangle R_{kl}. Obviously $\Delta R_{kl} = \Delta x_k \, \Delta y_l$.

Definition 1. The number

$$\sigma=\sum_{k=1}^{n}\sum_{l=1}^{p} f(\xi_k,\ \eta_l)\cdot\Delta R_{kl} \tag{2.1}$$

is called the integral sum *of a function $f(x, y)$ corresponding to a given subdivision T of a rectangle R and a given choice of intermediate points $(\xi_k,\ \eta_l)$ on the subrectangles of T.*

The diagonal $\sqrt{(\Delta x_k)^2+(\Delta y_l)^2}$ will be called the *diameter* of a rectangle R_{kl}. The symbol Δ will designate the largest of the diameters of all subrectangles R_{kl}.

Definition 2. The number I is said to be the limit of the integral sums (2.1) *as $\Delta\to 0$ if given any positive number ε we can find a positive number δ such that for $\Delta<\delta$, regardless of the choice of points $(\xi_k,\ \eta_l)$ on subrectangles R_{kl}, we have*

$$|\sigma-I|<\varepsilon.$$

Definition 3. A function $f(x, y)$ is said to be integrable (in the sense of Riemann) *on a rectangle R if there is a finite limit I of integral sums of $f(x, y)$ as $\Delta\to 0$.*

The limit I is called a double integral *of $f(x, y)$ over R and is designated by one of the following symbols:*

$$I=\iint_R f(x,\ y)\,dx\,dy=\iint_R f(M)\,d\sigma.$$

Remark. Just as for a single definite integral (see Section 10.1 of [1]), it can be established that *any function $f(x, y)$ integrable on a rectangle R is bounded on that rectangle.*

We are thus justified in considering only bounded functions $f(x, y)$ in what follows.

2.1.2. Existence of a double integral for the rectangle. Darboux's theory developed in Chapter 10 of [1] for the single definite integral can be extended in full to the case of the double integral in the rectangle R. In view of close analogy we restrict ourselves to a general outline of the arguments.

Let M_{kl} and m_{kl} be the supremum and infimum of a function $f(x, y)$ on a subrectangle R_{kl}. Form for the given subdivision T of R two sums

the upper one

$$S=\sum_{k=1}^{n}\sum_{l=1}^{p} M_{kl}\cdot\Delta R_{kl}$$

and the lower one

$$s=\sum_{k=1}^{n}\sum_{l=1}^{p} m_{kl}\cdot\Delta R_{kl}.$$

The following statements hold (their proofs are quite similar to those given in Section 10.2.2 of [1]).

1°. *For any fixed subdivision T and any $\varepsilon > 0$ we can choose intermediate points (ξ_k, η_l)* on *subrectangles R_{kl} so that the integral sum σ satisfies the inequalities $0 < S - \sigma < \varepsilon$.*

Points (ξ_k, η_l) can also be chosen in such a way that the integral sum satisfies the inequalities $0 \leqslant \sigma - s < \varepsilon$.

2°. *If a subdivision T' of a rectangle R is obtained by adding new straight lines to the straight lines generating the subdivision T, then the upper sum S' of T' is not greater than the upper sum S of T, and the lower sum s' of T' is not less than the lower sum s of T, i.e.*

$$s \leqslant s', \quad S' \leqslant S.$$

3°. *Let T' and T'' be any two subdivisions of a rectangle R. Then the lower sum of one of them does not exceed the upper sum of the other. That is, if s', S' and s'', S'' are respectively the lower and upper sums of T' and T'', then*

$$s' \leqslant S'', \quad s'' \leqslant S'.$$

4°. *A set $\{S\}$ of the upper sums of a given function $f(x, y)$ for all possible subdivisions of a rectangle R is bounded below. A set $\{s\}$ of the lower sums is bounded above.*

Thus there are numbers

$$\bar{I} = \inf\{S\}, \quad \underline{I} = \sup\{s\}$$

called respectively the **upper and lower Darboux integrals** *(of the function $f(x, y)$ over R).*

It can easily be seen that $\underline{I} \leqslant \bar{I}$.

5°. *Let T' be a subdivision of a rectangle R obtained from a subdivision T by adding to it p new straight lines and let s', S' and s, S be respectively the lower and upper sums of T' and T.*

Then for the differences $S - S'$ and $s' - s$ we can obtain an estimate dependent on the maximum diameter Δ of a subrectangle of the subdivision T, the number p of the straight lines added, the supremum and infimum M and m of the function $f(x, y)$ on R and on the diameter d of R.

That is,

$$S - S' \leqslant (M - m) \cdot p \cdot \Delta \cdot d,$$

$$s' - s \leqslant (M - m) \cdot p \cdot \Delta \cdot d.$$

6°. *The upper and lower Darboux integrals* $\bar{I}$ *and* $\underline{I}$ *of a function* $f(x, y)$ *over a rectangle* R *are respectively the limits of the upper and lower sums as* $\Delta \to 0$*.

The following main theorem results from properties 1° to 6°.

Theorem 2.1. For a function $f(x, y)$ *bounded on a rectangle* R *to be integrable on* R *it is necessary and sufficient that given any* $\varepsilon > 0$ *there should be a subdivision* T *of* R *for which* $S - s < \varepsilon$.

As in Chapter 10 of [1], Theorem 2.1 in conjunction with the theorem on the uniform continuity of a function allows us to distinguish major classes of integrable functions.

Theorem 2.2. Any function $f(x, y)$ *continuous in a rectangle* R *is integrable on* R.

Definition 1. We give the name of an elementary figure *to a set of points that are a sum of a finite number of rectangles (with sides parallel to the* x *and* y *axes)***.

Defenition 2. We say that the function $f(x, y)$ *has the* I-property *in a rectangle* R *(in an arbitrary closed domain* D*) if: (1)* $f(x, y)$ *is bounded in* R *(in* D*); (2) given any* $\varepsilon > 0$ *we can find an elementary figure that contains all the points and lines of discontinuity of* $f(x, y)$ *and has the area less than* ε.

Theorem 2.3. If a function $f(x, y)$ *has the I-property in a rectangle* R*, then it is integrable on* R.

The proofs of Theorems 2.2 and 2.3 are quite similar to those of Theorem 10.3 and Theorem 10.4 in [1].

2.1.3. Definition and existence of a double integral for an arbitrary domain. In Section 11.2.1 of [1] we introduced the *squarability and area* of a plane figure Q. These notion can be carried over without any modifications to the case of an arbitrary bounded set Q of points in the plane.

In all the definitions and statements of Section 11.2.1 of [1] we could take an arbitrary bounded set Q instead of the figure Q.

In that section we also gave a definition of a curve (or the boundary of a figure) of *zero area*: Γ is called a *curve of zero area* if given any $\varepsilon > 0$ we can find a polygon that contains all points of Γ and has an area less than ε.

---

* The concept of the limit of upper or lower sums is defined in full analogy with the notion of the limit of integral sums. That is, the number $\bar{I}$ is said to be *the limit of upper sums* S as $\Delta \to 0$ if given any $\varepsilon > 0$ we can find $\delta > 0$ such that $|S - \bar{I}| < \varepsilon$ for $\Delta < \delta$.

** Notice that a sum of a finite number of absolutely arbitrary rectangles (with sides parallel to the x and y axes) can be represented as a sum of an also finite number of rectangles *having no interior points in common* (with sides parallel to the above axes). In Definition 1 therefore we can take rectangles both with and without common interior points.

Note that in the definition "polygon" can be replaced by "elementary figure". This follows from the fact that any elementary figure is a polygon and any polygon with area less than ε is contained in an elementary figure having an area less than a number 8ε*.

It is easy to prove the following *statement.*

If Γ is of zero area and if the plane is covered with a square net with spacing h, then given any $\varepsilon > 0$ we can find $h > 0$ such that the sum of the areas of all the squares having points in common with Γ is less than ε.

Indeed, given any $\varepsilon > 0$ we can find some elementary figure Q containing Γ inside itself and having an area less than $\varepsilon/4$. After that it remains to notice that with the spacing of the square net h sufficiently small, all the squares having points in common with Γ are contained in the elementary figure resulting from replacing every rectangle Q by a twice as large rectangle with the same centre.

It should be stressed that the class of zero-area curves is very vast. Belonging to this class, for example, is any rectifiable curve (see Theorem 11.3 in [1]).

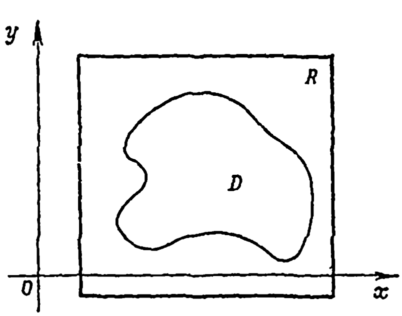


Fig. 2.3

We now proceed to define the double integral for an arbitrary two-dimensional domain D.

Let D be a closed bounded domain whose boundary Γ has zero area and let $f(x, y)$ be a function defined and bounded in D.

Denote by R any rectangle (with sides parallel to coordinate axes) containing the domain D (Fig. 2.3).

We define in R the following function:

$$F(x, y) = \begin{cases} f(x, y) & \text{at the points of } D, \\ 0 & \text{at the other points of } R. \end{cases} \tag{2.2}$$

Definition. *A function $f(x, y)$ will be said* to be integrable in the domain *D if the function $F(x, y)$ is integrable in the rectangle R.*

---

* Indeed, (1) a polygon is equal to a finite sum of triangles; (2) every triangle equals the sum (or difference) of two right-angled triangles; (3) a right-angled triangle is contained in a rectangle twice as large in area; (4) any rectangle equals the sum of a finite number of squares and one rectangle the ratio of whose sides is between 1 and 2; (5) any square is contained in a square twice as large in area and with sides parallel to the x and y axes; (6) any rectangle with ratio of its sides between 1 and 2 can be completed to a square and is therefore contained in a square four times as large in area and with sides parallel to the x and y axes.

The number $I = \iint\limits_R F(x, y)\,dx\,dy$ will be called a double integral of $f(x, y)$ over D and designated

$$I = \iint\limits_D f(x, y)\,dx\,dy = \iint\limits_D f(M)\,d\sigma.$$

Remark 1. From this definition it follows at once that $\iint\limits_D 1\cdot dx\,dy$ *is equal to the area of* D. Indeed, subjecting the corresponding rectangle R to smaller and smaller subdivisions we find that the upper sums of the subdivisions are equal to the areas of the elementary figures containing D and the lower sums to the areas of the elementary figures contained in D.

Remark 2. *Let the function* $f(x, y)$ *be integrable in a bounded squarable domain* D, *let the plane be covered by a square net with spacing* h, *let* $C_1, C_2, \ldots, C_{n(h)}$ *be the squares of the net contained entirely in* D, let (ξ_k, η_k) *be an arbitrary point of a square* C_l, *and* $m_k = \inf\limits_{C_k} f(x, y)$ $(k = 1, 2, \ldots, n(h))$. *Then each of the sums*

$$\sum_{k=1}^{n(h)} f(\xi_k, \eta_k)\cdot h^2, \quad \sum_{k=1}^{n(h)} m_l\cdot h^2$$

has a limit as $h \to 0$ *equal to* $\iint\limits_D f(x, y)\,dx\,dy$.

To prove this it suffices to notice that the sums differ from the ordinary integral sum (respectively from the lower sum) of a function $f(x, y)$ in D only in having no terms corresponding to squares with points in common with the boundary Γ of D, the sum of all the absent terms being less in absolute value than the product of the supremum M of the function $|f(x, y)|$ in D and the area S of the elementary figure consisting of the squares having common points with Γ. By the statement proved above $S \to 0$ as $h \to 0$.

Regarding the definition we have given, the question naturally arises as to whether the fact of the existence of a double integral and its value I depends (1) on the choice of coordinate axes Ox and Oy in the plane; (2) on the choice of a rectangle R on which we define the function $F(x, y)$.

In the next subsection we shall give another definition of the integrability of a function $f(x, y)$ and of the double integral that depends neither on the choice of coordinate axes nor on the choice of a rectangle R and prove the equivalence of the two definitions.

For the time being we shall demonstrate the following *main theorem* which is almost immediate from Theorem 2.3 and the above definition.

Theorem 2.4. If a function $f(x, y)$ has the I-property in a domain D, then it is integrable in D.

Proof. For such a function $f(x, y)$ the function $F(x, y)$ defined by formula (2.2) has the I-property in the rectangle R.

Indeed, $F(x, y)$ is bounded in R and all discontinuity points and lines of the function either coincide with the corresponding discontinuites of $f(x, y)$ or lie on the boundary Γ of D. Since Γ has zero area, the theorem is proved.

Corollary 1. If a function $f(x, y)$ is bounded in a domain D and has discontinuities only on a finite number of rectifiable curves in that domain, then $f(x, y)$ is integrable in D.

Corollary 2. If $f(x, y)$ is integrable in D and $g(x, y)$ is bounded and coincides with $f(x, y)$ everywhere in D except for the set of points of zero area, then $g(x, y)$ is also integrable in D.

2.1.4. Definition of a double integral by means of arbitrary subdivisions of a domain. In the subsection above we defined the double integral on the basis of dividing a domain by straight lines into a finite number of subrectangles. Here we shall give another definition of the double integral based on dividing a domain D by curves of zero area into a finite number of subdomains of arbitrary form and prove this definition to be equivalent to that given above.

Let D be a closed bounded domain with boundary Γ of zero area. Divide D by means of a finite number of arbitrary curves of zero area into a finite number r of (not necessarily connected!) closed subdomains $D_1, D_2, \ldots, D_r$.

Notice that every domain D_i is squarable, for its boundary has zero area (see Section 11.2 in [1]) and denote by ΔD_i the area of a subdomain D_i.

In every subdomain D_i choose a point $P_i(\xi_i, \eta_i)$.

Definition 1. The number

$$\tilde{\sigma} = \sum_{i=1}^{r} f(P_i)\cdot\Delta D_i \tag{2.3}$$

is called the **integral sum** *of a function $f(x, y)$ corresponding to a given subdivision of a domain D into subdomains D_i and to a given choice of intermediate points P_i in the subdomains.*

The diameter of a subdomain D_i is the supremum of the distances between any two points of the subdomain. The symbol $\tilde{\Delta}$ designates the largest of the diameters of the subdomains $D_1, D_2, \ldots, D_r$.

Definition 2. A number I is said **to be the limit of the integral sums** *(2.3) as $\tilde{\Delta} \to 0$ if given any positive number ε we can find a positive number δ such that for $\tilde{\Delta} < \delta$, regardless of the choice of points P_i, in subdomains D_i*

$$|\tilde{\sigma} - I| < \varepsilon.$$

Definition 3 (general definition of integrability).
A function $f(x, y)$ is said to be integrable (in the sense of Riemann) *in a domain D if there is a finite limit I of integrable sums $\tilde{\sigma}$ of $f(x, y)$ as $\tilde{\Delta} \to 0$. The limit I is called a* double integral of $f(x, y)$ over D.

We prove the following fundamental theorem.

Theorem 2.5. The above general definition of integrability is equivalent to that given in Section 2.1.3.

Proof. It is obvious that if a function $f(x, y)$ is integrable according to the general definition of integrability and its double integral by that definition is equal to I, then $f(x, y)$ is integrable according to the definition of Section 2.1.3 too and has by that definition the same double integral I.

It remains to prove that if $f(x, y)$ is integrable in D according to the definition of Section 2.1.3 and I is a double integral of $f(x, y)$ over D by that definition, then for the function $f(x, y)$ there is a limit of integral sums $\tilde{\sigma}$ equal to I as $\tilde{\Delta} \to 0$.

Denote by $\tilde{M}_i$ and $\tilde{m}_i$ the supremum and infimum of a function $f(x, y)$ in a subdomain D_i and consider the upper and lower sums

$$\tilde{S} = \sum_{i=1}^{r} \tilde{M}_i \cdot \Delta D_i \text{ and } \tilde{s} = \sum_{i=1}^{r} \tilde{m}_i \cdot \Delta D_i.$$

Since for any subdivision

$$\tilde{s} \leqslant \tilde{\sigma} \leqslant \tilde{S},$$

it is sufficient to prove that the two sums $\tilde{S}$ and $\tilde{s}$ converge to I as $\tilde{\Delta} \to 0$.

We wish to prove that given any $\varepsilon > 0$ we can find $\delta > 0$ such that each of the sums $\tilde{S}$ and $\tilde{s}$ deviates from I by less than ε as soon as $\tilde{\Delta} < \delta$.

Take an arbitrary $\varepsilon > 0$. For this ε we can find *a subdivision T* of a rectangle R containing the domain D into subrectangles R_k such that for that subdivision

$$S - s < \frac{\varepsilon}{2}. \tag{2.4}$$

Denote by M_0 the supremum $|f(x, y)|$ in D and enclose all the segments of the straight lines generating T and the boundary Γ of D in an elementary figure whose area is less than $\varepsilon/4M_0$.

Then automatically there is a *positive* supremum δ of the distance between two points one of which belongs to the boundary of the

elementary figure and the other to the segments of the straight lines generating the subdivision T or to the boundary Γ of the domain D*.

We prove that for the sums S and s of any subdivision of D satisfying the condition $\Delta < \delta$ we have

$$\tilde{S} < S + \frac{\varepsilon}{2}, \tag{2.5}$$

$$s - \frac{\varepsilon}{2} < \tilde{s}. \tag{2.6}$$

We restrict ourselves to proving inequality (2.5), for the proof of (2.6) is similar.

Remove from S all terms $\tilde{M}_i \cdot \Delta D_i$ corresponding to domains D_i, each *failing to be entirely in one subrectangle of the subdivision T*. All such subdomains D_i belong to the above elementary figure, so the grand total of the areas of such domains is less than $\varepsilon/4M_0$.

The sum of all removed terms $\tilde{M}_i \cdot \Delta D_i$ is therefore less than $\varepsilon/4$. Thus, within the error not greater than $\varepsilon/4$,

$$\tilde{S} = \sum{}' \tilde{M}_i \cdot \Delta D_i, \tag{2.7}$$

where the prime means that the sum is taken only over subdomains D_i *lying entirely in the corresponding rectangles of the subdivision T*.

Now we replace on the right of (2.7) the suprema $\tilde{M}_i$ in subdomains D_i contained in a subrectangle R_k by the supremum M_k in R_k. Then we get

$$\sum{}' \tilde{M}_i \cdot \Delta D_i \leqslant \sum_k M_k \cdot \Delta \tilde{R}_k, \tag{2.8}$$

where $\Delta \tilde{R}_k$ denotes the area of the domain $\tilde{R}_k$ equal to the sum of all subdomains D_i contained entirely in the subrectangle R_k.

All subdomains $R_k - \tilde{R}_k$ belong to the elementary figure chosen above. Consequently

$$\sum_k (\Delta R_k - \Delta \tilde{R}_k) < \frac{\varepsilon}{4M_0},$$

---

* Indeed, consider two sets: (1) a set $\{P\}$ of all the points of the elementary figure's boundary and (2) a set $\{Q\}$ of all the points of the segments of T and of the boundary Γ of D. Both sets $\{P\}$ and $\{Q\}$ are bounded and *closed*. Suppose the infimum δ of the distance ρ (P, Q) equals zero. Then we can find two sequences of points $\{P_n\}$ and $\{Q_n\}$ such that ρ $(P_n, Q_n) \to 0$. By the Bolzano-Weierstrass theorem we may choose convergent subsequences $\{P_{k_n}\}$ and $\{Q_{k_n}\}$ of $\{P_n\}$ and $\{Q_n\}$, whose limits P and Q (by closure) belong to $\{P\}$ and $\{Q\}$ respectively. But then ρ $(P, Q) = 0$, i.e. the points P and Q coincide, which is impossible, for the set $\{Q\}$ lies strictly inside the elementary figure and has no common points with $\{P\}$. The contradiction obtained proves that δ is positive.

and therefore

$$\left|S-\sum_k M_k\cdot\Delta\tilde{R}_k\right|=\left|\sum_k M_k(\Delta R_k-\Delta\tilde{R}_k)\right|<\frac{\varepsilon}{4}.$$

Thus, within the error not greater than $\varepsilon/4$,

$$\sum_k M_k\cdot\Delta\tilde{R}_k=S. \tag{2.9}$$

Comparing equations (2.7) and (2.9), which hold within the error not greater than $\varepsilon/4$, with (2.8) we obtain inequality (2.5).

The proof of inequality (2.6) is similar.

From (2.5) and (2.6) we get

$$s-\frac{\varepsilon}{2}<\tilde{s}\leqslant\tilde{S}<S+\frac{\varepsilon}{2}. \tag{2.10}$$

Since by (2.4) each of the sums s and S deviates from I by less than $\varepsilon/2$, each of the sums $\tilde{s}$ and $\tilde{S}$, by (2.10), deviates from I by less than ε. The theorem is proved.

2.2. BASIC PROPERTIES OF THE DOUBLE INTEGRAL

The properties of a double integral (and their derivation) are quite similar to the corresponding properties of a single definite integral. Therefore we restrict ourselves to formulating these properties.

1°. *Additivity*. If a function $f(x, y)$ is integrable in a domain D and if a curve Γ of zero area divides the domain D into two connected domains D_1 and D_2 having no interior points in common, then $f(x, y)$ is integrable in either of the domains D_1 and D_2, with

$$\iint\limits_D f(x, y)\,dx\,dy=\iint\limits_{D_1} f(x, y)\,dx\,dy+\iint\limits_{D_2} f(x, y)\,dx\,dy$$

2°. *Linear property*. If functions $f(x, y)$ and $g(x, y)$ are integrable in a domain D and α and β are any real numbers, then $[\alpha\cdot f(x, y)+\beta\cdot g(x, y)]$ is also integrable in D, with

$$\iint\limits_D [\alpha\cdot f(x, y)+\beta\cdot g(x, y)]\,dx\,dy=$$

$$=\alpha\iint\limits_D f(x, y)\,dx\,dy+\beta\iint\limits_D g(x, y)\,dx\,dy.$$

3°. If functions $f(x, y)$ and $g(x, y)$ are integrable in a domain D, then so is their product.

4°. If $f(x, y)$ and $g(x, y)$ are both integrable in a domain D and everywhere in D we have $f(x, y) \leqslant g(x, y)$, then

$$\iint_D f(x, y)\,dx\,dy \leqslant \iint_D g(x, y)\,dx\,dy.$$

5°. If $f(x, y)$ is integrable in D, then so is $|f(x, y)|$, with

$$\left|\iint_D f(x, y)\,dx\,dy\right| \leqslant \iint_D |f(x, y)|\,dx\,dy.$$

(Of course, the integrability of $|f(x, y)|$ in D does not imply that of $f(x, y)$ in D.)

6°. *Mean value theorem.* If both functions $f(x, y)$ and $g(x, y)$ are integrable in a domain D, the function $g(x, y)$ is nonnegative (nonpositive) everywhere in D, and M and m are the supremum and infimum of $f(x, y)$ in D, then we can find a number μ satisfying $m \leqslant \mu \leqslant M$ and such that we have the formula

$$\iint_D f(x, y)\,g(x, y)\,dx\,dy = \mu \iint_D g(x, y)\,dx\,dy. \tag{2.11}$$

In particular, if $f(x, y)$ is continuous in D and D *is bound*, then we can find in D* a point (ξ, η) such that $\mu = f(\xi, \eta)$ and formula (2.11) becomes

$$\iint_D f(x, y)\,g(x, y)\,dx\,dy = f(\xi, \eta) \iint_D g(x, y)\,dx\,dy.$$

7°. *Important geometrical property.* $\iint_D 1 \cdot dx\,dy$ is equal to the area of a domain D. (As was already noted above, this property follows immediately from the definition of integrability in Section 2.1.3.)

2.3. REDUCING A DOUBLE INTEGRAL TO AN ITERATED SINGLE INTEGRAL

Reduction of a double integral to an iterated single integral to be discussed in this section is one of the efficient ways of computing a double integral.

2.3.1. The case of the rectangle.

Theorem 2.6. Let there be a double integral $\iint_R f(x, y)\,dx\,dy$ *for a function* $f(x, y)$ *in a rectangle* $R = [a \leqslant x \leqslant b] \times [c \leqslant y \leqslant d]$.

* By Theorem 14.5 of [1].

Also let there be for each x of the closed interval $a \leqslant x \leqslant b$ a single integral

$$I(x) = \int_c^d f(x, y)\, dy. \tag{2.12}$$

Then there is an iterated integral

$$\int_a^b I(x)\, dx = \int_a^b dx \int_c^d f(x, y)\, dy \tag{2.13}$$

and we have

$$\iint_R f(x, y)\, dx\, dy = \int_a^b dx \int_c^d f(x, y)\, dy.$$

Proof. As in Section 2.1, divide the rectangle R by means of the points $a = x_0 < x_1 < x_2 < \ldots < x_n = b$ and $c = y_0 < y_1 < < y_2 < \ldots < y_p = d$ into $n \cdot p$ subrectangles

$R_{kl} = [x_{k-1} \leqslant x \leqslant x_k] \times [y_{l-1} \leqslant y \leqslant y_l]$

$(k = 1, 2, \ldots, n;\ l = 1, 2, \ldots, p)$.

Set $\Delta x_k = x_k - x_{k-1}$, $\Delta y_l = y_l - y_{l-1}$ and denote by M_{kl} and m_{kl} the supremum and infimum of $f(x, y)$ on a subrectangle R_{kl}. Everywhere on R_{kl} then

$$m_{kl} \leqslant f(x, y) \leqslant M_{kl}. \tag{2.14}$$

Set $x = \xi_k$, where ξ_k is an arbitrary point in a closed interval $[x_{k-1}, x_k]$, and after that integrate (2.14) for y going from y_{l-1} to y_l. We get

$$m_{kl} \cdot \Delta y_l \leqslant \int_{y_{l-1}}^{y_l} f(\xi_k, y)\, dy \leqslant M_{kl} \cdot \Delta y_l. \tag{2.15}$$

Summing (2.15) over all l from 1 to p and using notation (2.12) we have

$$\sum_{l=1}^{p} m_{kl} \cdot \Delta y_l \leqslant I(\xi_k) \leqslant \sum_{l=1}^{p} M_{kl} \cdot \Delta y_l. \tag{2.16}$$

Next we multiply (2.16) by Δx_k and sum over all k from 1 to n. We get

$$\sum_{k=1}^{n} \sum_{l=1}^{p} m_{kl} \Delta x_k \Delta y_l \leqslant \sum_{k=1}^{n} I(\xi_k) \cdot \Delta x_k \leqslant \sum_{k=1}^{n} \sum_{l=1}^{p} M_{kl} \cdot \Delta x_k \cdot \Delta y_l. \tag{2.17}$$

Let the greatest diameter Δ of subrectangles tend to zero. Then so does the greatest of the lengths Δx_h. The framing terms in (2.17), the lower and the upper sum, then converge to a double integral $\iint\limits_R f(x, y)\,dx\,dy$.

There is therefore a limit on the middle term too in (2.17), equal to the same double integral. But that limit by the definition of the single integral equals

$$\int_a^b I(x)\,dx = \int_a^b dx \int_c^d f(x, y)\,dy.$$

This proves the existence of the iterated integral and equation (2.13). Thus the theorem is proved.

Remark. In Theorem 2.6 we may interchange x and y, i.e. we may assume the existence of the double integral and the existence for any y in the closed interval $c \leqslant y \leqslant d$ of the single integral

$$K(y) = \int_a^b f(x, y)\,dx.$$

Then the theorem will state the existence of the iterated integral

$$\int_c^d K(y)\,dy = \int_c^d dy \int_a^b f(x, y)\,dx$$

and the equation

$$\iint\limits_R f(x, y)\,dx\,dy = \int_c^d dy \int_a^b f(x, y)\,dx. \tag{2.18}$$

2.3.2. The case of an arbitrary domain.

Theorem 2.7. Let the following conditions hold: (1) *a domain D is bounded, closed and such that any straight line parallel to the Oy axis intersects the boundary of that domain in at most two points whose ordinates are $y_1(x)$ and $y_2(x)$, where $y_1(x) \leqslant y_2(x)$* (Fig. 2.4); (2) *the function $f(x, y)$ allows the existence of a double integral*

$$\iint\limits_D f(x, y)\,dx\,dy$$

and the existence for any x of a single integral

$$\int_{y_1(x)}^{y_2(x)} f(x, y)\,dy.$$

Under these conditions there is an iterated integral

$$\int_{x_1}^{x_2} dx \int_{y_1(x)}^{y_2(x)} f(x,\ y)\, dy$$

(x_1 and x_2 being the smallest and largest abscissas of the points of D) and we have

$$\iint_D f(x,\ y)\, dx\, dy = \int_{x_1}^{x_2} dx \int_{y_1(x)}^{y_2(x)} f(x,\ y)\, dy. \tag{2.19}$$

Proof. Denote by R a rectangle with sides parallel to the coordinate axes and containing a domain D, and by $F(x,\ y)$ a function coinciding with $f(x,\ y)$ at the points of D and equal to zero at the other

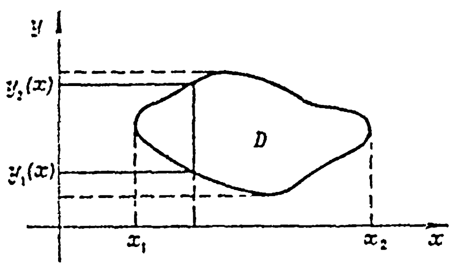


Fig. 2.4

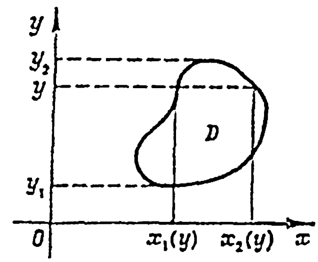


Fig. 2.5

points of R. For the function $F(x,\ y)$ all hypotheses of Theorem 2.7 hold in R, and therefore we have formula (2.13) which (considering that $F(x,\ y)$ is zero outside D and coincides with $f(x,\ y)$ in D) goes over into (2.19). The theorem is proved.

Remark 1. In Theorem 2.7 we may interchange x and y, i.e. we may assume that the following two conditions hold: (1) the domain D is such that any straight line parallel to the x axis intersects the boundary of the domain D in at most two points whose abscissas are $x_1(y)$ and $x_2(y)$, where $x_1(y) \leqslant x_2(y)$ (Fig. 2.5); (2) $f(x,\ y)$ allows the existence over D of a double integral and the existence for any y of a single integral

$$\int_{x_1(y)}^{x_2(y)} f(x,\ y)\, dx.$$

If the two conditions hold there is an iterated integral

$$\int_{y_1}^{y_2} dy \int_{x_1(y)}^{x_2(y)} f(x,\ y)\, dx$$

(y_1 and y_2 being the smallest and largest ordinates of the points of D) and

$$\iint\limits_D f(x, y)\,dx\,dy = \int\limits_{y_1}^{y_2} dy \int\limits_{x_1(y)}^{x_2(y)} f(x, y)\,dx. \tag{2.19'}$$

Example. Let a domain D be the circle $x^2 + y^2 \leqslant R^2$ (Fig. 2.6) and $f(x, y) = x^2 (R^2 - y^2)^{3/2}$. Any straight line parallel to the x

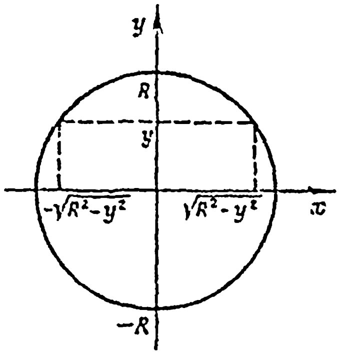


Fig. 2.6

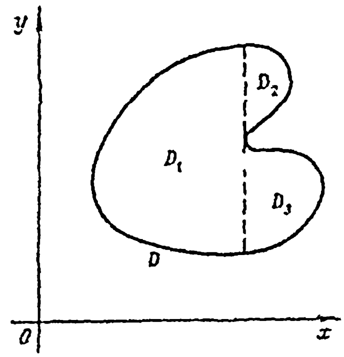


Fig. 2.7

axis intersects the boundary of D in at most two points whose abscissas are $x_1 = -\sqrt{R^2 - y^2}$ and $x_2 = \sqrt{R^2 - y^2}$ (see Fig. 2.6). Therefore, applying formula (2.19') we get

$$\iint\limits_D f(x, y)\,dx\,dy = \int\limits_{-R}^{R} dy \int\limits_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} x^2 (R^2 - y^2)^{3/2}\,dx =$$

$$= \int\limits_{-R}^{R} (R^2 - y^2)^{3/2} \left[\int\limits_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} x^2\,dx \right] dy = \frac{2}{3} \int\limits_{-R}^{R} (R^2 - y^2)^3\,dy = \frac{64}{105} R^7.$$

Remark 2. In case a domain D fails to satisfy the requirements of Theorem 2.7 or Remark 1 to it, it is often possible to divide that domain into a sum of a finite number of domains of this type that have no interior points in common. Then the integral over D is, by additivity (see property 1° of Section 2.2), equal to the sum of the integrals over the corresponding domains. Thus it is possible to divide the domain D depicted in Fig. 2.7 into the sum of three domains, D_1, D_2, and D_3, to each of which we can apply either Theorem 2.7 or Remark 1.

2.4. TRIPLE INTEGRALS AND n-FOLD MULTIPLE INTEGRALS

The theory of double integral we have presented can be extended, without introducing any new ideas or complications, to the case of the *triple integral* or to *n-fold multiple* integral in general. We shall discuss the main points of the theory of n-fold multiple integral.

First of all, let us define the volume of an n-dimensional rectangular parallelepiped to be equal to the product of the lengths of all of its edges emanating from a single vertex.

Let us further give the name of an *elementary body* to a set of points of an n-dimensional space that is a sum of a finite number of n-dimensional rectangular parallelepipeds without common interior points and with edges parallel to the coordinate axes.

The volume of any elementary body is known to us and is equal to the sum of the volumes of the constituent parallelepipeds.

Now let D be a bounded domain in n-dimensional Euclidean space. The *lower volume* of D is the supremum $\underline{V}$ of the volumes of all elementary bodies contained in D, and the *upper volume* of D is the infimum $\overline{V}$ of all elementary bodies containing T.

It can easily be seen that $\underline{V} \leqslant \overline{V}$*.

A domain D is said to be *cubable* if $\underline{V} = \overline{V}$. The number $V = \underline{V} = \overline{V}$ is the *n-dimensional volume* of D.

The following statement can be proved in complete analogy with the case of the plane domain.

For an n-dimensional domain to be cubable, it is necessary and sufficient that for any positive number ε *there should be two elementary bodies, one containing* D *and the other contained in* D*, the difference between whose volumes in absolute value is less than* ε.

Let us agree to apply the term *surface* (or *manifold*) *of n-dimensional volume zero* to a closed set all the points of which lie in an elementary body of arbitrarily small n-dimensional volume.

It is obvious that an n-dimensional domain D is cubable if and only if the boundary of the domain is a manifold of n-dimensional volume zero.

First an n-fold multiple integral of a function of n variables $f(x_1, x_2, \ldots, x_n)$ is defined in an n-dimensional rectangular parallelepiped R whose edges are parallel to the coordinate axes.

To this end we divide each of n edges of R into a finite number of segments to obtain a subdivision T of R into a finite number of n-dimensional subparallelepipeds**.

---

* The inequality $\underline{V} \leqslant \overline{V}$ can be proved in exactly the same way as $\underline{P} \leqslant \overline{P}$ was in Section 11.2.1 of [1].

** It may be said that a subdivision T is carried out by means of a finite number of $(n-1)$-dimensional hyperplanes parallel to the coordinate axes.

For the subdivision T the integral, upper, and lower sums of any bounded function $f(x_1, x_2, \ldots, x_n)$ are defined in complete analogy with the case $n = 2$.

The n-fold multiple integral of $f(x_1, x_2, \ldots, x_n)$ over the parallelepiped R is defined to be the limit of integral sums as the length of the largest of the diagonals of the n-dimensional subparallelepipeds tends to zero.

As for the case $n = 2$, Darboux's theory establishes a necessary and sufficient condition of integrability in the following form: *for a function f to be integrable in a parallelepiped R it is necessary and sufficient that given any $\varepsilon > 0$ there should be a subdivision T of R for which the difference between the upper and the lower sum should be less than* ε.

After that it is easy to define the n-fold multiple integral of f over an arbitrary closed bounded n-dimensional domain D with boundary of n-dimensional volume zero.

This integral can be defined as an integral over a D-containing n-dimensional rectangular parallelepiped R (with edges parallel to the coordinate axes) of a function F coinciding with f in D and zero outside D.

It is natural to designate an n-fold multiple integral of a function $f(x_1, x_2, \ldots, x_n)$ over a domain D by the symbol

$$\iint\limits_D \cdots \int f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n. \tag{2.20}$$

To abbreviate notation where no confusion may arise, however, we shall denote the integral (2.20) briefly

$$\int\limits_D f(x)\, dx. \tag{2.20'}$$

In (2.20′), by x we mean a point $x = (x_1, x_2, \ldots, x_n)$ in E^n, by dx a product $dx = dx_1\, dx_2 \ldots dx_n$* and by $\int\limits_D$ an n-fold multiple integral over an n-dimensional domain D.

Just as for the case $n = 2$, we can prove integrability over an n-dimensional domain D of any function f possessing in D the *I-property* (i.e. a D-bounded function all of whose discontinuity points lie in an elementary body of arbitrarily small n-dimensional volume). In general no change in an integrable function f on a set of points of n-dimensional volume zero would change the value of the integral of that function.

To define the n-fold multiple integral we can use division of a domain D into a finite number of subdomains of arbitrary form by

* This product is usually called an element of volume in E^n.

means of a finite number of arbitrary manifolds of volume zero. Repeating the arguments for Theorem 2.5 we prove that such a general definition of the n-fold multiple integral is equivalent to the above definition.

In complete analogy with Theorem 2.6 and Theorem 2.7 we establish the *iterated integration formula* for the integral (2.20).

Let an n-dimensional domain D_n possess the property that any straight line parallel to the axis Ox_1 intersects its boundary at at most two points whose projections onto the axis Ox_1 are

$$a(x_2, x_3, \ldots, x_n) \text{ and } b(x_2, x_3, \ldots, x_n),$$

where $a(x_2, x_3, \ldots, x_n) \leqslant b(x_2, x_3, \ldots, x_n)$.

Further let the function $f(x_1, x_2, \ldots, x_n)$ allow the existence of an n-fold multiple integral

$$\int\int_{D_n} \ldots \int f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2 \ldots dx_n$$

and the existence for any $x_2, x_3, \ldots, x_n$ of a single integral

$$\int_{a(x_2, x_3, \ldots, x_n)}^{b(x_2, x_3, \ldots, x_n)} f(x_1, x_2, \ldots, x_n)\,dx_1.$$

Then there is an $(n-1)$-fold multiple integral

$$\int\int_{D_{n-1}} \ldots \int dx_2\,dx_3 \ldots dx_n \int_{a(x_2, x_3, \ldots, x_n)}^{b(x_2, x_3, \ldots, x_n)} f(x_1, x_2, \ldots, x_n)\,dx_1$$

over an $(n-1)$-dimensional domain D_{n-1} which is a projection of D_n onto the coordinate hyperplane $Ox_2x_3, \ldots, x_n$, and we have the iterated integration formula

$$\int\int_{D_n} \ldots \int f(x_1, x_2, \ldots x_n)\,dx_1\,dx_2 \ldots dx_n =$$

$$= \int\int_{D_{n-1}} \ldots \int dx_2\,dx_3 \ldots dx_n \int_{a(x_2, x_3, \ldots, x_n)}^{b(x_2, x_3, \ldots, x_n)} f(x_1, x_2, \ldots, x_n)\,dx_1. \tag{2.21}$$

Of course, in the statement above any of the variables $x_2, x_3, \ldots, x_n$ may also play the part of x_1.

We shall agree to say that a domain D is *simple* if every straight line parallel to any coordinate axis either intersects the boundary of the domain in at most two points or has an entire segment on that boundary.

Given a simple domain we can use the iterated integration formula for any of the variables $x_1, x_2, \ldots, x_n$.

An n-dimensional rectangular parallelepiped (whose edges are not necessarily parallel to the coordinate axes) may exemplify a simple domain.

In conclusion we note that properties 1° to 7° formulated in Section 2.2 for the case of the double integral remain valid for the n-fold multiple integral.

In particular $\int\int \underset{D}{\ldots} \int 1 \cdot dx_1\, dx_2 \ldots dx_n$ is equal to the n-dimensional volume $V(D)$ of a domain D.

In addition, as for the case $n = 2$, the following *statement* holds.

Let $f(x_1, x_2, \ldots, x_n)$ be a function integrable over a bounded cubable domain D. Also let E^n be a space covered by a net of n-dimensional cubes with edge h; let $C_1, C_2, \ldots, C_{n(h)}$ be those cubes of the net that are contained entirely in D; let $(\xi_1^{(h)}, \xi_2^{(h)}, \ldots, \xi_n^{(h)})$ be an arbitrary point of a cube C_h; and let m_h be the infimum of f in a cube C_k $(k = 1, 2, \ldots, n(h))$. Then the sums

$$\sum_{k=1}^{n(h)} f(\xi_1^{(k)}, \xi_2^{(h)}, \ldots, \xi_n^{(h)}) \cdot h^n \text{ and } \sum_{k=1}^{n(h)} m_k \cdot h^n$$

have a limit, as $h \to 0$, equal to $\int\int \underset{D}{\ldots} \int f(x_1, x_2, \ldots, x_n)\, dx_1 \times$ $\times\, dx_2 \ldots dx_n$.

2.5. CHANGE OF VARIABLES IN AN n-FOLD MULTIPLE INTEGRAL

The purpose of the present section is to justify the formula for change of variables in an n-fold multiple integral.

The formula to be established is one of the major tools for computing an n-fold multiple integral.

Suppose that a function $f(y_1, y_2, \ldots, y_n)$ allows the existence of the n-fold multiple integral

$$\int_D f(y)\, dy = \int\int \underset{D}{\ldots} \int f(y_1, y_2, \ldots, y_n)\, dy_1\, dy_2 \ldots dy_n \tag{2.22}$$

over some bounded closed cubable domain D in the space of variables $y_1, y_2, \ldots, y_n$. Further suppose that we change from $y_1, y_2, \ldots, y_n$ to some new variables $x_1, x_2, \ldots, x_n$, i.e. make a transformation

$$\begin{cases} y_1 = \psi_1(x_1, x_2, \ldots, x_n) \\ y_2 = \psi_2(x_1, x_2, \ldots, x_n) \\ \ldots\ldots\ldots\ldots\ldots \\ y_n = \psi_n(x_1, x_2, \ldots, x_n). \end{cases} \tag{2.23}$$

In shorthand, (2.23) will be denoted by the symbol

$y = \psi(x)$,

where x and y are points of the n-dimensional space $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ and ψ is the collection of n functions $\psi_1, \psi_2, \ldots, \psi_n$.

We denote by D' the domain in the space of $x_1, x_2, \ldots, x_n$ transforming under (2.23) to D, i.e. set $D = \psi(D')$*.

We prove that if the functions (2.23) have continuous partial derivatives of the first order in D' and if the Jacobian

$$\frac{\mathscr{D}(y)}{\mathscr{D}(x)} = \frac{\mathscr{D}(y_1, y_2, \ldots, y_n)}{\mathscr{D}(x_1, x_2, \ldots, x_n)} \tag{2.24}$$

is nonzero in D', then for the integral (2.22) the following *change of variables formula* is true:

$$\int\limits_D f(y)\,dy = \int\limits_{D'} f[\psi(x)] \left| \frac{\mathscr{D}(y)}{\mathscr{D}(x)} \right| dx. \tag{2.25}$$

In explicit notation (2.25) has the form

$$\int\int\limits_D \ldots \int f(y_1, y_2, \ldots, y_n)\,dy_1\,dy_2 \ldots dy_n =$$

$$= \int\int\limits_{D'} \ldots \int f[\psi_1(x_1, \ldots, x_n), \ldots, \psi_n(x_1, \ldots, x_n)] \times$$

$$\times \left| \frac{\mathscr{D}(y_1, \ldots, y_n)}{\mathscr{D}(x_1, \ldots, x_n)} \right| dx_1 \ldots dx_n. \tag{2.25'}$$

Thus we shall prove the following *main* theorem.

Theorem 2.8. If the transformation (2.23) *converts the domain D' into D and is one-to-one and if the functions* (2.23) *have in D' continuous partial derivatives of the first order and the nonzero Jacobian* (2.24)**, *then provided the integral* (2.22) *exists* (2.25') *is true.*

The proof of Theorem 2.8 is not elementary. The basic idea behind it is that we first justify formula (2.25) for the case where the transformation (2.23) is *linear* and then reduce the general transformation (2.23) to that case.

For convenience we divide the proof into several steps.

Proof of Theorem 2.8.

1°. *Lemma 1. If a transformation $z = \psi(x)$ is a superposition*

* We assume that the transformation (2.23) has the inverse and that $D' = \psi^{-1}(D)$.

** Note that when the hypotheses of Theorem 2.8 hold equations (2.23) can be solved for $x_1, x_2, \ldots, x_n$, the inverse transformation $x = \psi^{-1}(y)$ obtained in this way having in D, by Theorem 14.2 of [1], continuous partial derivatives of the first order and a nonzero Jacobian $\frac{\mathscr{D}(x)}{\mathscr{D}(y)}$.

(or, as it is usually called, a **product**) *of two transformations* $y = \psi_1(x)$ *and* $z = \psi_2(y)$, *then the Jacobian* $\frac{\mathcal{D}(z)}{\mathcal{D}(x)}$ *taken at any point* $\overset{0}{x} = (\overset{0}{x}_1, \overset{0}{x}_2, \ldots, \overset{0}{x}_n)$ *is the product of the Jacobians* $\frac{\mathcal{D}(y)}{\mathcal{D}(x)}$ *and* $\frac{\mathcal{D}(z)}{\mathcal{D}(y)}$ *taken respectively at* $\overset{0}{x} = (\overset{0}{x}_1, \overset{0}{x}_2, \ldots, \overset{0}{x}_n)$ *and* $\overset{0}{y} = (\overset{0}{y}_1, \overset{0}{y}_2, \ldots, \overset{0}{y}_n)$, *where* $\overset{0}{y} = \psi_1(\overset{0}{x})$, *i. e.*

$$\frac{\mathcal{D}(z)}{\mathcal{D}(x)} = \frac{\mathcal{D}(z)}{\mathcal{D}(y)} \cdot \frac{\mathcal{D}(y)}{\mathcal{D}(x)}. \tag{2.26}$$

In explicit notation formula (2.26) assumes the form

$$\frac{\mathcal{D}(z_1, z_2, \ldots, z_n)}{\mathcal{D}(x_1, x_2, \ldots, x_n)} = \frac{\mathcal{D}(z_1, z_2, \ldots, z_n)}{\mathcal{D}(y_1, y_2, \ldots, y_n)} \cdot \frac{\mathcal{D}(y_1, y_2, \ldots, y_n)}{\mathcal{D}(x_1, x_2, \ldots, x_n)}. \tag{2.26'}$$

Proof of Lemma 1. The element at the intersection of the ith row and the kth column of the Jacobian $\frac{D(z)}{D(x)}$ is equal to $\frac{\partial z_i}{\partial x_k}$, with the partial derivative taken at a point $\overset{0}{x}$. By the indirect differentiation rule (see Section 14.7 of [1])

$$\frac{\partial z_i}{\partial x_k} = \sum_{l=1}^{n} \frac{\partial z_i}{\partial y_l} \cdot \frac{\partial y_l}{\partial x_k}, \tag{2.27}$$

all partial derivatives $\frac{\partial y_l}{\partial x_k}$ on the right of (2.27) are taken at $\overset{0}{x}$ and all partial derivatives $\frac{\partial z_i}{\partial y_l}$ at the corresponding points $\overset{0}{y} = \psi_1(\overset{0}{x})$.

Equations (2.27), true for any $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n$, and the theorem on the determinant of a product of two matrices (see "Linear Algebra") lead directly to formula (2.26).

The proof of Lemma 1 is complete.

2°. Before stating the next lemma we recall the definition of *linear transformation of coordinates.*

Linear transformation *is a transformation of the form*

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n, \end{cases} \tag{2.28}$$

where a_{ik} ($i = 1, 2, \ldots, n$; $k = 1, 2, \ldots, n$) are arbitrary constant numbers.

The linear transformation (2.28) will be briefly denoted by $y = Tx$, where x and y are points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ of a space E^n and T is a matrix $T = \| a_{ik} \|$ $(i = 1, 2, \ldots, n;\ k = 1, 2, \ldots, n)$.

The matrix T is usually called the **matrix of linear transformation.**

If the determinant of the matrix of linear transformation $\det T$ is nonzero, then $y = Tx$ is said to be a nonsingular linear transformation. For such a transformation, by the Cramer theorem*, we can solve equations (2.28) for $x_1, x_2, \ldots, x_n$ and assert the existence of the inverse transformation $x = T^{-1}y$ which is also linear and nonsingular.

Further notice that for the linear transformation (2.28) the Jacobian $\frac{\mathscr{D}(y)}{\mathscr{D}(x)}$ coincides with the determinant of the matrix T of the transformation, i. e.

$$\frac{\mathscr{D}(y)}{\mathscr{D}(x)} = \det T. \tag{2.29}$$

The main purpose of this and the next two subsections is to prove that the change of variable formula (2.25) is valid for an arbitrary linear nonsingular transformation (2.28). In view of relation (2.29) it suffices to prove that for any linear nonsingular transformation $y = Tx$ we have the formula

$$\int_{D} f(y)\,dy = \int_{T^{-1}D} f(Tx)\,|\det T|\,dx \tag{2.30}$$

(provided there is an integral on the left of the formula).

In the present subsection we shall prove that (2.30) holds for two special types of linear transformations: (1) linear transformation T_i^{λ} consisting in multiplying the ith coordinate by a real number $\lambda \neq 0$, with the other coordinates remaining unchanged**, and (2) linear transformation T_{ij} consisting in adding to the ith coordinate the jth coordinate, all the other coordinates except the ith coordinate remaining unchanged***.

Lemma 2. If $f(y)$ is integrable in D, then (2.30) *holds for either of the transformations T_i^{λ} and T_{ij}.*

Proof of Lemma 2. Denote by R an n-dimensional rectangular parallelepiped containing D and by F a function equal to f in D and to zero in $R - D$. It suffices to prove that for either of the

---

* See "Linear Algebra" for the Cramer theorem.

** In symbols

$$(x_1, x_2, \ldots, x_n) \to (x_1, \ldots, x_{i-1}, \lambda x_i, x_{i+1}, \ldots, x_n).$$

*** In symbols

$$(x_1, x_2, \ldots, x_n) \to (x_1, \ldots, x_{i-1}, x_i + x_j, x_{i+1}, \ldots, x_n).$$

transformations T_i^λ and T_{ij} we have the formula

$$\int_R F(y)\,dy = \int_{T^{-1}R} F(Tx)\cdot|\det T|\,dx, \tag{2.31}$$

in which T denotes one of the transformations, T_λ^i or T_{ij}.

An elementary calculation shows that

$$\det T_i^\lambda = \lambda, \quad \det T_{ij} = 1. \tag{2.32}$$

Moreover, it is obvious that if R is a rectangular parallelepiped $a_k \leqslant y_k \leqslant b_k$ $(k = 1, 2, \ldots, n)$, then $[T_i^\lambda]^{-1} R$ is a rectangular parallelepiped

$$\begin{cases} a_k \leqslant x_k \leqslant b_k \text{ when } k \neq i, \\ \dfrac{a_i}{\lambda} \leqslant x_i \leqslant \dfrac{b_i}{\lambda}, \end{cases} \tag{2.33}$$

and $[T_{ij}]^{-1} R$ is a fortiori a cubable domain

$$\begin{cases} a_k \leqslant x_k \leqslant b_k \text{ when } k \neq i, \\ a_i - x_j \leqslant x_i \leqslant b_i - x_j. \end{cases} \tag{2.34}$$

On the basis of the iterated integration formula (2.21)

$$\int_R F(y)\,dy = \int_{a_1}^{b_1} \ldots \int_{a_{i-1}}^{b_{i-1}} \int_{a_{i+1}}^{b_{i+1}} \ldots \int_{a_n}^{b_n} dy_1 \ldots dy_{i-1}\,dy_{i+1} \ldots$$

$$\ldots dy_n \int_{a_i}^{b_i} F(y_1, \ldots, y_n)\,dy_i. \tag{2.35}$$

Applying to a single integral with respect to a variable y_i the change of variable formula $y_i = \lambda x_i$ for the case of T_i^λ and $y_i = x_i + x_j$ for the case of T_{ij} (see Section 10.7 in [1]) we get:

(a)

$$\int_{a_i}^{b_i} F(y_1, \ldots, y_n)\,dy_i =$$

$$= \begin{cases} \displaystyle\int_{a_i/\lambda}^{b_i/\lambda} F(y_1, \ldots, y_{i-1}, \lambda x_i, y_{i+1}, \ldots, y_n)\,\lambda\,dx_i \text{ when } \lambda > 0, \\ \displaystyle\int_{b_i/\lambda}^{a_i/\lambda} F(y_1, \ldots, y_{i-1}, \lambda x_i, y_{i+1}, \ldots, y_n)\,(-\lambda)\,dx_i \text{ when } \lambda < 0; \end{cases} \tag{2.36}$$

for the case of T_i^{λ}, and

(b)

$$\int_{a_i}^{b_i} F(y_1, \ldots, y_n)\, dy_i =$$

$$= \int_{a_i - x_j}^{b_i - x_j} F(y_1, \ldots, y_{i-1}, x_i + x_j, y_{i+1}, \ldots, y_n)\, dx_i \qquad (2.37)$$

for T_{ij}.

Inserting (2.36) into (2.35), applying once again formula (2.21) and taking into account the equation $y_k = x_k$, with $k \neq i$, the form (2.33) of $[T_i^{\lambda}]^{-1} R$ and the first of the equations (2.32) we obtain formula (2.31) for the case of T_i^{λ}.

Similarly, inserting (2.37) into (2.35), applying the iterated integration formula and taking into account the equations $y_k = x_k$, with $k \neq i$, the form (2.34) of $[T_{ij}]^{-1} R$ and the second of the equations (2.32) we obtain formula (2.31) for the case of T_{ij}. This completes the proof of Lemma 2.

3°. *Lemma 3. Any nonsingular linear transformation T can be represented as a superposition of a finite number of linear transformations of the types T_i^{λ} and T_{ij}.*

Proof of Lemma 3. We first verify that the linear transformation T' consisting in interchanging some two coordinates can be represented as a superposition of six transformations of the types T_i^{λ} and T_{ij}. Indeed, let T' consist in interchanging the ith and jth coordinates (the other coordinates remaining unchanged). Then it is easy to verify that*

$$T' = T_i^{-1} T_{ij} T_j^{-1} T_{ji} T_i^{-1} T_{ij}. \qquad (2.38)$$

Now notice that by a finite number of interchanges of two rows and two columns a quite arbitrary linear nonsingular transformation T can be reduced to a linear transformation (2.28) with matrix $\| a_{ik} \|$ all of whose so-called *principal minors* are nonzero, i.e. all determinants

$$\Delta_k = \begin{vmatrix} a_{11} & \ldots & a_{1k} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ a_{k1} & \ldots & a_{kk} \end{vmatrix} \quad (k = 1, 2, \ldots, n). \qquad (2.39)$$

* Indeed, keeping in writing only the ith and jth coordinates we get, on performing the chain of transformations (2.38), $(x_i, x_j) \to (x_i + x_j, x_j) \to (-x_i - x_j, x_j) \to (-x_i - x_j, -x_i) \to (-x_i - x_j, x_i) \to (-x_j, x_i) \to (x_j, x_i)$.

It remains to prove that the last linear transformation can be represented as a superposition of a finite number of transformations of the types T_i^λ and T_{ij}.

We shall prove this by induction.

Since $\Delta_1 = a_{11} \neq 0$, using $T_1^{a_{11}}$ we obtain $(x_1, x_2, \ldots, x_n) \to$ $\to (a_{11}x_1, x_2, \ldots, x_n)$.

Now suppose that by superposing a finite number of transformations of the types T_i^λ and T_{ij} we have succeeded in reducing the original sequence of coordinates $(x_1, x_2, \ldots, x_n)$ to the form

$$(a_{11}x_1 + \ldots + a_{1k}x_k, \ldots, a_{k1}x_1 + \ldots + a_{kk}x_k, x_{k+1}, \ldots, x_n). \quad (2.40)$$

To complete the induction it suffices to prove that by superposing a finite number of transformations of the types T_i^λ and T_{ij} we can reduce the sequence of coordinates (2.40) to the form

$$\begin{aligned}(a_{11}x_1 + \ldots + a_{1(k+1)}x_{k+1}, \ldots, a_{k1}x_1 + \ldots + a_{k(k+1)}x_{k+1},\\ a_{(k+1)1}x_1 + \ldots + a_{(k+1)(k+1)}x_{k+1}, x_{k+2}, \ldots, x_n).\end{aligned} \quad (2.41)$$

First for every i for which the element $a_{i(k+1)}$ is nonzero we perform successively a pair of transformations $T_{i(k+1)}T_{k+1}^{a_{i(k+1)}}$ (for those i for which $a_{i(k+1)} = 0$ the corresponding pair of transformations is not performed). Superposition of all the pairs of transformations reduces the sequence (2.40) to the form

$$\begin{aligned}(a_{11}x_1 + \ldots + a_{1(k+1)}x_{k+1}, \ldots, a_{k1}x_1 + \ldots +\\ + a_{k(k+1)}x_{k+1}, x_{k+1}, x_{k+2}, \ldots, x_n).\end{aligned} \quad (2.42)$$

Further notice that since the minor (2.39) is nonzero, so is the determinant

$$\begin{vmatrix} a_{11} & \ldots & a_{1k} & a_{1(k+1)} \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1} & \ldots & a_{kk} & a_{k(k+1)} \\ 0 & \ldots & 0 & 1 \end{vmatrix} \quad (2.43)$$

equal to it.

But then there are real numbers $\lambda_1, \ldots, \lambda_k, \lambda_{k+1}$ such that a linear combination of the rows of (2.43) with those numbers is equal to*

$$a_{(k+1)1}, \ldots, a_{(k+1)k}, a_{(k+1)(k+1)}. \quad (2.44)$$

This means that if for every $j = 1, 2, \ldots, k+1$ for which $\lambda_j \neq 0$ we perform successively a pair of transformations $T_{(k+1)j}\, T_j^{\lambda_j}$

* To prove this it suffices to add to the matrix of the determinant (2.43) the row (2.44) and to apply the theorem on the principal minor (see "Linear Algebra").

(for those j for which $\lambda_j = 0$ the corresponding pair of transformations is not performed,) then superposition of all the pairs of transformations made converts the sequence (2.42) into (2.41). This completes the induction and the proof of Lemma 3.

4°. *Lemma 4. For an arbitrary linear nonsingular transformation* (2.28) *the change of variables formula* (2.30) *holds, provided the integral on the left of* (2.30) *exists.*

To prove Lemma 4 it suffices to notice that formula (2.30) holds for each of the transformations of the types T_i^λ and T_{ij} (Lemma 2) and that an arbitrary linear nonsingular transformation (2.28) can be represented as superposition of a finite number of transformations of the types T_i^λ and T_{ij} (Lemma 3), superposition of linear transformations involving multiplication of the corresponding Jacobians (Lemma 1).

Corollary of Lemma 4. If G is an arbitrary cubable domain in E^n, T is an arbitrary nonsingular linear transformation, then the n-dimensional volume $V(G)$ of G and the n-dimensional volume $V(TG)$ of the image TG of G are related by

$$V(TG) = |\det T| \cdot V(G). \tag{2.45}$$

To prove this it is enough to put in (2.30) $f \equiv 1$, $D = TG$ and take into account the fact that $T^{-1}D = G$.

5°. Now we proceed to justify the change of variables formula (2.25) for a quite arbitrary transformation $y = \psi(x)$ satisfying the hypotheses of Theorem 2.8.

It should be emphasized that when the hypotheses of Theorem 2.8 hold the integrals on the left and on the right of (2.25) both exist, so that we are to prove only the equality of the integrals.

Let us agree to denote by $J_{ij}(x)$ the elements of the Jacobi matrix $\frac{\partial \psi_i}{\partial x_j}$ $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, n)$ taken at a point $x = (x_1, x_2, \ldots, x_n)$.

The Jacobi matrix itself, $\| J_{ij}(x) \|$ will be denoted by $J_\psi(x)$.

It is convenient to introduce the *norm of a point* $x = (x_1, x_2, \ldots, x_n)$ and the *norm of a matrix* $A = \| a_{ij} \|$ $(i = 1, 2, \ldots, n;$ $j = 1, 2, \ldots, n)$.

The norm of a point $x = (x_1, x_2, \ldots, x_n)$ *is a number denoted by the symbol* $\| x \|$ *and equal to* $\max\limits_{i=1,2,\ldots,n} |x_i|$.

The norm of a matrix $A = \| a_{ij} \|$ *is a number denoted by the symbol* $\| A \|$ *and equal to* $\max\limits_{i=1,2,\ldots,n} \left[\sum\limits_{j=1}^{n} |a_{ij}| \right]$.

Notice that with such a definition of the norms of a point and of a matrix the equation $y = Ax$ implies that

$$\| y \| \leqslant \| A \| \cdot \| x \|. \tag{2.46}$$

In addition it is easy to verify that for the unit matrix E we have $\| E \| = 1$.

In this subsection we shall prove the following lemma.

Lemma 5. If the hypotheses of Theorem 2.8 hold and if C is an n-dimensional cube in D', then the n-dimensional volumes of the cube C and of its image $\psi(C)$ are related by

$$V(\psi(C)) \leqslant [\max_{x \in C} \| J_\psi(x) \|]^n \cdot V(C). \tag{2.47}$$

Proof. Let C be an n-dimensional cube with centre at a point $\overset{0}{x} = (\overset{0}{x}_1, \overset{0}{x}_2, \ldots, \overset{0}{x}_n)$ and with edge $2s$. Then the cube C can be defined by

$$\| x - \overset{0}{x} \| \leqslant s. \tag{2.48}$$

By the Taylor formula for a function of n variables $\psi_i(x)$ (see Section 14.5.3 in [1]) there is a number θ_i in $0 < \theta_i < 1$ such that

$$\psi_i(x) - \psi_i(\overset{0}{x}) = \sum_{j=1}^{n} J_{ij}(\overset{0}{x} + \theta(x - \overset{0}{x}))(x_j - \overset{0}{x}_j).$$

From this and from relation (2.46) we conclude that

$$\| \psi(x) - \psi(\overset{0}{x}) \| \leqslant \max_{x \in C} \| J_\psi(x) \| \cdot \| x - \overset{0}{x} \|. \tag{2.49}$$

Setting $y = \psi(x)$, $\overset{0}{y} = \psi(\overset{0}{x})$, we get from (2.49) and (2.48)

$$\| y - \overset{0}{y} \| \leqslant s \cdot \max_{x \in C} \| J_\psi(x) \|.$$

Thus *when a point x is changed inside an n-dimensional cube C with edge $2s$ the image of x remains inside an n-dimensional cube whose edge is $2s \cdot \max\limits_{x \in C} \| J_\psi(x) \|$.*

This leads at once to the cubability of the image $\psi(G)$ of any cubable set G* (in particular, to the cubability of $\psi(C)$) and to inequality (2.47). This completes the proof of Lemma 5.

6°. *Lemma 6. Let the hypotheses of Theorem 2.8 hold and let G be a cubable subset of D'. Then for an n-dimensional volume of the image $\psi(G)$ of G we have***

$$V(\psi(G)) \leqslant \int_G |\det J_\psi(x)| \, dx. \tag{2.50}$$

* Indeed, the boundary of any cubable set G is a set of n-dimensional volume zero, and such a set, by the statement proved, can be transformed into a set whose n-dimensional volume is also zero.

** The fact itself of the cubability of $\psi(G)$ follows from the statement proved in the preceding lemma.

Proof of Lemma 6. We first prove that for any non-singular linear transformation T and for any n-dimensional cube C contained in D'

$$V(\psi(C)) \leqslant |\det T| \cdot [\max_{x \in C} \| T^{-1} J_\psi(x) \|]^n \cdot V(C). \qquad (2.51)$$

By the corollary of Lemma 4, for any cubable set G and for any linear transformation T^{-1}

$$V(T^{-1}G) = |\det T^{-1}| \cdot V(G).$$

Thus if $G = \psi(C)$, then*

$$V(\psi(C)) = |\det T| \cdot V(T^{-1}\psi(C)). \qquad (2.52)$$

We evaluate the right-hand side of (2.52) using (2.47) by taking (2.47) not for the transformation ψ but for the superposition of transformations $T^{-1}\psi$. Then

$$V(\psi(C)) \leqslant |\det T| \cdot [\max_{x \in C} \| J_{T^{-1}\psi}(x) \|]^n \cdot V(C). \qquad (2.53)$$

Considering that the Jacobian matrix of a linear transformation coincides with the matrix of that transformation we get by Lemma 1

$$J_{T^{-1}\psi}(x) = T^{-1} J_\psi(x).$$

But this exactly means that inequality (2.53) may be rewritten as (2.51).

This proves inequality (2.51).

Now, to prove Lemma 6, we cover the space E^n with a net of n-dimensional cubes with edge h, and let $C_1, C_2, \ldots, C_{n(h)}$ be those of the cubes that are contained entirely in G, and let G_h denote the sum of all the cubes.

On choosing in every cube C_i a point x_i we write for that cube inequality (2.51) setting $T = J_\psi(x_i)$. We get

$$V(\psi(C_i)) \leqslant |\det J_\psi(x_i)| \{\max_{x \in C_i} \| [J_\psi(x_i)]^{-1} \cdot J_\psi(x) \|\}^n \cdot V(C_i).$$

Summing the last inequality over all i from 1 to $n(h)$ we get

$$V(\psi(G_h)) \leqslant \sum_{i=1}^{n(h)} |\det J_\psi(x_i)| \cdot \{\max_{x \in C_i} \| [J_\psi(x_i)]^{-1} \cdot J_\psi(x) \|\}^n \cdot V(C_i). \qquad (2.54)$$

Since the elements of the Jacobian matrix $J_\psi(x)$ are continuous functions of a point x in the whole of D' and all the more so in G and the product $[J_\psi(x)]^{-1} \cdot J_\psi(x)$ is a unit matrix whose norm is equal to unity, we have

$$\lim_{h \to 0} \max_{x \in C_i} \| [J_\psi(x_i)]^{-1} \cdot J_\psi(x) \| = 1.$$

* We take into account the fact that $T \cdot T^{-1} = E$, so that $\det T \cdot \det T^{-1} = 1$.

But then it follows from the statement formulated at the end of Section 2.4 that the limit of the entire right-hand side of (2.54) as $h \to 0$ exists and is $\int\limits_G |\det J_\psi(x)|\,dx$.

That same statement implies that $\lim\limits_{h\to 0} G_h = G$, so that in the limit as $h \to 0$ we obtain from (2.54) inequality (2.50). The proof of Lemma 6 is complete.

7°. *Lemma 7. Let the hypotheses of Theorem 2.8 all hold and suppose in addition that $f(y)$ is* nonnegative *in D. Then the change of variables formula* (2.25) *is valid.*

Proof. Cover E^n with a net of n-dimensional cubes with edges h and let $C_1, C_2, \ldots, C_{n(h)}$ be those of the cubes that are contained entirely in D. Further let $G_i = \psi^{-1}(C_i)$. Writing for every domain G_i inequality (2.50) we have

$$V(C_i) \leqslant \int\limits_{G_i} |\det J_\psi(x)|\,dx. \tag{2.55}$$

Now let m_i be the infimum of the function $f(y)$ on a cube C_i (or equivalently the infimum of $f[\psi(x)]$ in G_i). Multiplying both sides of (2.55) by m_i and summing over all i from 1 to $n(h)$ we have

$$\sum_{i=1}^{n(h)} m_i V(C_i) \leqslant \sum_{i=1}^{n(h)} m_i \int\limits_{G_i} |\det J_\psi(x)|\,dx. \tag{2.56}$$

By the statement at the end of Section 2.4. the left-hand side of (2.56) has a limit as $h \to 0$ equal to $\int\limits_D f(y)\,dy$. Since the sum of all domains G_i is contained in D'* and f is *nonnegative*, the right-hand side of (2.56) for any $h > 0$ does not exceed

$$\int\limits_{D'} f[\psi(x)]\cdot|\det J_\psi(x)|\,dx.$$

Thus we obtain in the limit as $h \to 0$ from (2.56)

$$\int\limits_D f(y)\,dy \leqslant \int\limits_{D'} f[\psi(x)]\cdot|\det J_\psi(x)|\,dx. \tag{2.57}$$

In our arguments above we may interchange D and D' and consider $g(x) = f[\psi(x)]\cdot|\det J_\psi(x)|$ in D' instead of $f(y)$ in D. Using

---

* In view of the fact that $\sum\limits_{i=1}^{n(h)} C_i$ is contained in D, $D' = \psi^{-1}(D)$, $G_i = \psi^{-1}(C_i)$.

Lemma 1 and the theorem on the determinant of the product of two matrices we obtain an inequality that has an opposite sense

$$\int_{D'} f[\psi(x)]\cdot|\det J_\psi(x)|\,dx \leqslant \int_{D} f(y)\,dy. \tag{2.58}$$

From (2.57) and (2.58) we obtain the change of variables formula (2.25). The proof of Lemma 7 is complete.

8°. It remains for us to complete the proof of Theorem 2.8, i.e. to get rid of the additional requirement on the nonnegativity of the function $f(y)$ imposed in Lemma 7.

Let $f(y)$ be a quite arbitrary function integrable over D and let M be the supremum of $|f(y)|$ in D*.

By Lemma 7 the change of variables formula (2.25) is valid for either of the nonnegative functions $f_1(y) \equiv M$ and $f_2(y) = M - f(y)$.

But then the linearity of the integral implies the validity of formula (2.25) for the difference $f_1(y) - f_2(y) = f(y)$ too. The proof of Theorem 2.8 is complete.

Remark 1. Under the hypotheses of Theorem 2.8 we may assume the Jacobian (2.24) to vanish on some set of points S of n-dimensional volume zero in D'. Indeed, S lies inside an elementary figure C of arbitrary small area, with by what was proved above

$$\int_{\psi(D'-C)} f(y)\,dy = \int_{D'-C} f[\psi(x)]\cdot|\det J_\psi(x)|\,dx. \tag{2.59}$$

Proceeding in formula (2.59) to the limit with respect to a sequence of elementary figures $\{C_k\}$ whose n-dimensional volume $V(C_k)$ tends to zero we see that (2.25) is valid also for the case under consideration.

Remark 2. Since the integral

$$I = \int\int_{D} \ldots \int 1\,dy_1\,dy_2 \ldots dy_n \tag{2.60}$$

is equal to the n-dimensional volume $V(D)$ of D, it is natural to call $dy_1\,dy_2 \ldots dy_n$ an *element of volume* in the Cartesian coordinate system $Oy_1y_2 \ldots y_n$ considered.

Using the transformations (2.23) we transform from the Cartesian coordinates $y_1, y_2, \ldots, y_n$ to some new, generally curvilinear, coordinates $x_1, x_2, \ldots, x_n$. Since under such a transformation (by (2.25)) the integral (2.60) becomes

$$I = \int\int_{D} \ldots \int \left|\frac{\mathcal{D}(y_1, y_2, \ldots, y_n)}{\mathcal{D}(x_1, x_2, \ldots, x_n)}\right| dx_1\,dx_2 \ldots dx_n,$$

* Recall that the integrability of $f(y)$ in D implies the boundedness of $f(y)$ in D and the existence of the infimum and supremum.

it is natural to call

$$\left|\frac{D(y_1, y_2, \ldots, y_n)}{D(x_1, x_2, \ldots, x_n)}\right| dx_1\, dx_2 \ldots dx_n$$

an *element of volume* in curvilinear coordinate system $x_1, x_2, \ldots, x_n$.

Therefore the absolute value of the Jacobian is characterized by the "expansion" (or "shrinking") of volume under the transformation from Cartesian coordinates $y_1, y_2, \ldots, y_n$ to curvilinear coordinates $x_1, x_2, \ldots, x_n$.

Let us compute the element of volume in spherical and cylindrical coordinates.

1°. For the spherical coordinates (in three-dimensional space)

$$\begin{cases} x = r\cos\varphi\sin\theta, \\ y = r\sin\varphi\sin\theta, \quad (r \geqslant 0,\ 0 \leqslant \theta \leqslant \pi,\ 0 \leqslant \varphi < 2\pi) \\ z = r\cos\theta \end{cases}$$

the Jacobian has the form

$$\frac{D(x, y, z)}{D(r, \varphi, \theta)} = \begin{vmatrix} \cos\varphi\sin\theta & -r\sin\varphi\sin\theta & r\cos\varphi\cos\theta \\ \sin\varphi\sin\theta & r\cos\varphi\sin\theta & r\sin\varphi\cos\theta \\ \cos\theta & 0 & -r\sin\theta \end{vmatrix} = r^2\sin\theta.$$

Therefore the element of volume equals $r^2 \sin\theta\, dr\, d\theta\, d\varphi$.

2°. For the cylindrical coordinates (in three-space)

$$\begin{cases} x = r\cos\varphi, \\ y = r\sin\varphi, \quad (r \geqslant 0,\ 0 \leqslant \varphi < 2\pi) \\ z = z \end{cases}$$

the Jacobian has the form

$$\frac{D(x, y, z)}{D(r, \varphi, z)} = \begin{vmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{vmatrix} = r.$$

Therefore the volume element is $r\, dr\, d\varphi\, dz$.

In particular, for polar coordinates in the plane the element of area is $r\, dr\, d\varphi$.

3°. In an n-dimensional space spherical coordinates are defined by the equations*

$$\begin{cases} x_1 = r \sin \theta_1 \sin \theta_2 \ldots \sin \theta_{n-1}, \\ x_m = r \cos \theta_{m-1} \prod\limits_{k=m}^{n-1} \sin \theta_k \quad \text{when } m = 2, 3, \ldots, n-1, \\ x_n = r \cos \theta_{n-1}, \end{cases}$$

where the spherical radius r and the spherical angles $\theta_1, \theta_2, \ldots, \theta_{n-1}$ vary over the range $r \geqslant 0$, $0 \leqslant \theta_1 < 2\pi$, $0 \leqslant \theta_m \leqslant \pi$ for $m = 2, 3, \ldots, n-1$.

We can see that in this case the Jacobian has the form

$$\frac{\mathscr{D}(x_1, x_2, \ldots, x_n)}{\mathscr{D}(r, \theta_1, \ldots, \theta_{n-1})} = r^{n-1} \prod_{k=1}^{n-1} \sin^{k-1} \theta_k.$$

Thus the volume element in n-dimensional spherical coordinates is $r^{n-1}\, dr \prod\limits_{k=1}^{n-1} \sin^{k-1} \theta_k \, d\theta_k$.

Example 1. Find the volume of a body bounded by the surface

$$(x^2 + y^2 + z^2)^2 = a^3 z, \tag{2.61}$$

where $a > 0$.

The body is symmetrical with respect to the coordinate planes Oyz and Oxz and lies above the plane Oxy. It is sufficient therefore to find the volume of the quarter of the body lying in the first octant.

Transforming to spherical coordinates we reduce equation (2.61) to the form

$$r = a \sqrt[3]{\cos \theta}.$$

Since in the first octant we have

$$0 \leqslant \theta \leqslant \frac{\pi}{2}, \quad 0 \leqslant \varphi \leqslant \frac{\pi}{2},$$

in view of the expression for the volume element in spherical coordinates the desired volume V is

$$V = 4 \int_0^{\pi/2} d\varphi \int_0^{\pi/2} d\theta \int_0^{a\sqrt[3]{\cos\theta}} r^2 \sin \theta \, dr.$$

---

* The reverse formulas expressing n-dimensional spherical coordinates in terms of Cartesian ones are of the form

$$r = \sqrt{x_1^2 + \ldots + x_n^2}, \quad \sin \theta_m = \frac{r_m}{r_{m+1}}, \quad \cos \theta_m = \frac{x_{m+1}}{r_{m+1}},$$

where $r_m = \sqrt{x_1^2 + \ldots + x_m^2}$, $m = 1, 2, \ldots, n-1$.

Thus

$$V=\frac{2\pi}{3}a^3\int_0^{\pi/2}\sin\theta\cos\theta\,d\theta=\frac{\pi a^3}{3}.$$

Example 2. Find the area of a figure bounded by the curve

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=\frac{x}{h}+\frac{y}{k} \tag{2.62}$$

(where $h>0$, $k>0$, $a>0$, $b>0$).

To find this area it is convenient to transform to the so-called *generalized polar coordinates*

$$\begin{cases} x=ar\cos\varphi, \\ y=br\sin\varphi \end{cases} \quad (0\leqslant\varphi\leqslant2\pi).$$

Equation (2.62) takes the form

$$r=\frac{a}{h}\cos\varphi+\frac{b}{k}\sin\varphi. \tag{2.63}$$

Since the left-hand side of (2.63) is nonnegative one should take only those values of φ for which the right-hand side of (2.63) is nonnegative.

On multiplying and dividing the right-hand side of (2.63) by $\sqrt{\frac{a^2}{h^2}+\frac{b^2}{k^2}}$ and determining φ_0 from the relations

$$\sin\varphi_0=\frac{a/h}{\sqrt{\frac{a^2}{h^2}+\frac{b^2}{k^2}}},\quad \cos\varphi_0=\frac{b/k}{\sqrt{\frac{a^2}{h^2}+\frac{b^2}{k^2}}}$$

we reduce (2.63) to the form

$$r=\sqrt{\frac{a^2}{h^2}+\frac{b^2}{k^2}}\sin(\varphi+\varphi_0). \tag{2.63'}$$

From the nonnegativity condition of the right-hand side of (2.63′) we find that $0\leqslant\varphi+\varphi_0\leqslant\pi$, i.e. $-\varphi_0\leqslant\varphi\leqslant\pi-\varphi_0$.

Considering that the Jacobian $\frac{\mathscr{D}(x,y)}{\mathscr{D}(r,\varphi)}$ equals abr, we obtain for the desired area S the following expression:

$$S=\int_{-\varphi_0}^{\pi-\varphi_0}d\varphi\int_0^{\sqrt{\frac{a^2}{h^2}+\frac{b^2}{k^2}}\sin(\varphi+\varphi_0)}abr\,dr=$$

$$=\frac{ab}{2}\left(\frac{a^2}{h^2}+\frac{b^2}{k^2}\right)\int_{-\varphi_0}^{\pi-\varphi_0}\sin^2(\varphi+\varphi_0)\,d\varphi=\frac{ab\pi}{4}\left(\frac{a^2}{h^2}+\frac{b^2}{k^2}\right).$$

We remark in conclusion that in computing a series of areas it is convenient to use a somewhat more general form of generalized polar coordinates

$$\begin{cases} x = ar\cos^{\alpha}\varphi, \\ y = br\sin^{\alpha}\varphi. \end{cases}$$

It is easy to see that for these coordinates

$$\frac{\mathscr{D}(x, y)}{\mathscr{D}(r, \varphi)} = \alpha abr\cos^{\alpha-1}\varphi\sin^{\alpha-1}\varphi.$$

SUPPLEMENT

ON THE APPROXIMATE CALCULATION OF n-Fold MULTIPLE INTEGRALS

We shall discuss the question of approximating the n-fold multiple integral

$$\int\int_{G_n}\cdots\int f(x_1, x_2, \ldots, x_n)\,dx_1\,dx_2\ldots dx_n \tag{2.64}$$

over some domain G_n of E^n, assuming first this domain to be an n-dimensional cube.

Supposing the integral (2.64) exists we consider the question of optimal ways of numerical integration.

The question has two aspects: (1) constructing numerical integration formulas optimal *for given classes of functions*; (2) constructing numerical integration formulas optimal *for each particular function of a given class.*

Let us consider each of these aspects separately.

2S.1. Formulas for numerical integration that are optimal for classes of functions. Let G_n be a unit n-dimensional cube $0 \leqslant x_k \leqslant 1$, $k = 1, 2, \ldots, n$.

We shall say that $f(x_1, \ldots, x_n)$ belongs in G_n to the *class* $D_n^{\alpha}(M)$ (respectively to the class $H_n^{\alpha}(M)$) if, provided all the derivatives appearing below exist, we have the inequalities

$$\left|\frac{\partial^{\beta} f}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}}\right| \leqslant M,$$

where

$$\beta = \sum_{k=1}^{n}\alpha_k \leqslant \alpha n, \quad \alpha_k \leqslant \alpha n$$

(respectively $\alpha_k \leqslant \alpha$).

We shall use the term *cubature formula* for an expression of the form

$$\int\int_{G_n}\cdots\int f(x_1, \ldots, x_n)\,dx_1\ldots dx_n = l_N(f) + R_N(f, l_N), \tag{2.65}$$

where

$$l_N(f) = \sum_{k=1}^{N} C_k f(x_1^{(k)}, \ldots, x_n^{(k)}).$$

The points $(x_1^{(k)}, \ldots, x_n^{(k)})$ will be called *nodes* and the numbers C_k are the *weights* of a given cubature formula, $R_N(f, l_N)$ being its *error*.

Our aim is to construct cubature formulas with error estimation accurate in order with respect to an infinitesimal $1/N$, where N is the number of nodes in the cubature formula.

N.S. Bakhvalov showed* that neither for classes $D_n^\alpha(M)$ nor for classes $H_n^\alpha(M)$ one can construct a cubature formula (2.65) with error estimation $R_N(f, l_N)$ better than $C(\alpha, n)\cdot M\cdot N^{-\alpha}$, where $C(\alpha, n)$ is some constant depending on α and n.

For classes $H_n^\alpha(M)$ the indicated estimation is obtained (in order with respect to $1/N$) if we take as l_N the product of one-dimensional quadrature formulas accurate for algebraic polynomials of degree $\alpha n - 1$.

Assuming that the number of nodes N is $N = m^n$, where m is an integer, we may put

$$l_N = \sum_{k_1=1}^{m} \cdots \sum_{k_n=1}^{m} C_{k_1} \cdots C_{k_n} f(x_{k_1}, \ldots, x_{k_n}), \tag{2.66}$$

where $\{x_{k_\nu}, C_{k_\nu}\}$, $\nu = 1, 2, \ldots, n$, are the nodes and weights of a one-dimensional quadrature formula accurate for algebraic polynomials**.

For the error of the cubature formula with l_N defined by equation (2.66) we have an *asymptotic* estimation (i.e., an estimation correct for sufficiently large values of N)

$$R_N(f, l_N) \approx \frac{C_1(\alpha, n)\, M}{N^\alpha}, \tag{2.67}$$

where $C_1(\alpha, n)$ is some constant depending on α and n.

For classes $H_n^\alpha(M)$ there is also a cubature formula close in order of magnitude of error to an optimal one. Such a formula is the number-theoretic formula of N.M. Korobov***.

$$l_N = \frac{1}{N}\sum_{k=1}^{N} f\left[\tau_\alpha\left(\frac{ka_1}{N}\right), \ldots, \tau_\alpha\left(\frac{ka_n}{N}\right)\right]\tau'_\alpha\left(\frac{ka_1}{N}\right)\cdots\tau'_\alpha\left(\frac{ka_n}{N}\right), \tag{2.68}$$

where $a_1, \ldots, a_n$ are integers, the so-called *optimal coefficients modulo N*, and $\tau_\alpha(x)$ are some special polynomials of degree $\alpha + 1$. For the error of the cubature formula with l_N defined by equation (2.68) we have

$$|R_N(f, l_N)| \leqslant \frac{C_2(\alpha, n)\, M}{N^\alpha} \ln^\beta N \tag{2.69}$$

---

* N.S. Bakhvalov. On approximating multiple integrals. Vestnik MGU, seria matematiki, fiziki, astronomii, No. 4 (1959), pp. 3-18.

** Such are, for example, the so-called Gauss formula or the Newton-Cotes formula (see, for example, "Methods of Computation" by I.S. Beryozin and N.P. Zhidkov).

*** N.M. Korobov. Number-theoretic methods in approximate analysis. Moscow, Fizmatgiz, 1963.

($C_2(\alpha, n)$ and β are constants depending only on α and n). The estimation (2.69) differs from the estimation best in order only by the multiplier $\ln^{\beta} N$.

Thus there are sufficiently good cubature formulas for each of the classes $D_n^{\alpha}(M)$ and $H_n^{\alpha}(M)$.

Of course, in practical applications of these formulas one should consider their advantages and drawbacks emerging in particular situations. Thus it should be remembered that in computing integrals with the aid of formula (2.66) the number of nodes N is not arbitrary but is equal to m^n. For example, for $n = 10$ and the function $f(x_1, \ldots, x_n)$ behaving roughly "equally" in all directions a reasonable number of nodes will be $N = 2^{10} = 1024$. When a greater accuracy is desirable, one can take the number of nodes equal to $N_1 = 3^{10} = 59{,}049$, but this will result in an almost 60-fold increase in computational work.

It should also be taken into account that for a "small" or "intermediate" number of nodes N the error of a cubature formula obtained using (2.66) may greatly differ from the right-hand side of (2.67)*.

On the other hand, it is more advantageous to use formula (2.66) when computing large series of integrals or integrals of functions containing expressions depending on a smaller number of variables than n.

Cubature formulas obtained using (2.68) are free of the disadvantage relating to the choice of the number of nodes N. It is appropriate to use these formulas for insufficiently smooth functions f and for a large value of the number of variables n (beginning with $n = 10$). It should be noted, however, that for the error of a cubature formula obtained using (2.68) one cannot point out a principal term similar to the one on the right-hand side of (2.67). This makes difficult both estimating errors in making computations and predicting the number of nodes N required for a given accuracy to be obtained.

28.2. On formulas for numerical integration optimal for every particular function. We note at the outset that the matter is involved and that little detailed work has been done on it.

To begin with, we revise the statement of the question under study. Suppose that a given function $f(x_1, x_2, \ldots, x_n)$ belongs to some class A_n and that we are given a set of methods of numerical integration $\{p_N\}$ of that function f. We shall seek in the given set a numerical integration method p_N^* such that its error $R_N(f, p_N^*)$ is the infimum of errors $R_N(f, p_N)$ of the set $\{p_N\}$ of the given function f.

In other words, we are seeking the best cubature formula for a given particular function f, not for the whole class A_n to which that function belongs**.

Take as a class A_n a set of functions infinitely differentiable everywhere in an entire cube G_n, except possibly some surface S of dimension $k < n$ on which those functions may go into infinity as $1/r_{xy}^{\gamma}$, where r_{xy} is the distance between a point $x = (x_1, \ldots, x_n)$ and a point on the surface $y = (y_1, \ldots, y_n)$, and $\gamma < n - k - 1$.

The set $\{p_N\}$ is defined as follows.

For every cubature formula σ_m exact for algebraic polynomials of degree $m - 1$ we define an element p_N of $\{p_N\}$ as a cubature formula obtained by

* Thus when using for (2.66) the Newton-Cotes quadrature formula the right-hand side of (2.67) is close to its left-hand side beginning with $N = (\alpha_n)^n$ (for instance, for $\alpha = 1$ and $n = 10$, beginning with $N = 10^{10}$), while when the Gauss formula is used for (2.66) the right-hand side of (2.67) is close to the left-hand side beginning with $N = (\alpha n/2)^n$ (i.e. beginning with $N \approx 10^7$ for $\alpha = 1$ and $n = 10$). Thus in constructing cubature formulas with l_N defined by (2.66) the Gauss formula is preferred to the Newton-Cotes formula.

** A formula best for a class of functions is roughly speaking best for the "worst" function of that class.

dividing the entire cube G_n into rectangular parallelepipeds and using for every such parallelepiped a formula σ_m with the proviso that the total number of nodes in the entire cube G_n should be N.

It is natural to expect the nodes of the cubature formula obtained in this way to be optimally allocated provided the error for each parallelepiped is constant.

At the computing centre of Moscow State University routines were compiled for computing double and triple integrals that realized automatic subdivision of domains of integration. They were based on a pair of cubature formulas σ_m and σ_{m_1} for $m_1 > m$.

The number $\rho = |\sigma_m - \sigma_{m_1}|$ was taken as the error estimation of a formula σ_m.

If ε is a given computational accuracy, then when $\rho \leqslant \varepsilon$ (for an entire cube G_n) we take as an approximate value of the integral the one defined by the formula σ_{m_1}, and when $\rho > \varepsilon$ the cube is divided into 2^n parts and the process is repeated all over again for each of the parts.

This method gives good results for computing double and triple integrals. As the number of measurements n increases, however, the application of the method runs into substantial difficulties arising from the fact that with m and m_1 fixed increasing n drastically increases the complexity of σ_m and σ_{m_1} and with m and m_1 decreased increasing n drastically increases the number of subdivisions.

In conclusion we note that when computing an n-fold multiple integral not over an n-dimensional cube G_n but over an arbitrary domain in E^n one should first make a transformation converting the domain into an n-dimensional cube. Besides, there are cubature formulas for some domains of special form (the ball, sphere, etc.)*.

2S.3. Example of approximate calculation of a multiple integral. Consider computing the fourfold integral

$$F(R, L, H) = \int_0^R r\,dr \int_0^L \rho\,d\rho \int_0^{2\pi} d\psi \int_0^{2\pi} [H + \rho^2 + r^2 - 2\rho r\cos(\varphi - \psi)]^{-3/2}\,d\varphi$$

to some accuracy ε for the values of parameters

$R = 1;\ 1.25;\ 1.5;\ 1.75;\ 2;\ 2.25;\ 2.5;\ 2.75;\ 3;\ L = 0.8;\ H = 1.$

A change of variables that maps the domain of integration into a unit cube reduces the integral to the form

$$F(R, L, H) = (2\pi)^2 \cdot R^2 \cdot L^2 \int_0^1 \int_0^1 \int_0^1 \int_0^1 [H^2 + L^2\rho^2 + R^2r^2 -$$

$$-2RL\rho r\cos 2\pi(\varphi - \psi)]^{-3/2}\, r\rho\,d\rho\,dr\,d\psi\,d\varphi.$$

The integrand is smooth. It is natural therefore to use in computing the integral a cubature formula based on (2.66). For each of the variables r and ρ it is natural to take the Gauss formula (a one-dimensional formula exact for algebraic polynomials), while for the variables φ and ψ it is better to take the trapezoidal formula (see Chapter 12 in [1]), for the integrand is periodic in each of these variables, and for periodic functions the trapezoidal formula gives the best results.

* Thus cubature formulas for the sphere have been studied in the works of S.L. Sobolev and his disciples.

Thus we get

$F(R, L, H) =$

$$= \left(\frac{2\pi RL}{m}\right)^2 \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \sum_{k_3=1}^{m_3} \sum_{k_4=1}^{m_4} C_{k_1} C_{k_2} x_{k_1} x_{k_2} \times$$

$$\times \left[H^2 + L^2 x_{k_2}^2 + R^2 x_{k_1}^2 - 2LRx_{k_1} x_{k_2} \cos 2\pi \frac{k_3 - k_4}{m} \right]^{-3/2}$$

here, (x_{k_ν}), (C_{k_ν}) are the nodes and weights of the corresponding quadrature formula).

To choose the values of m, m_1, and m_2 guaranteeing a required accuracy debugging calculations are made, successively increasing the number of nodes and comparing the results obtained.

CHAPTER 3

IMPROPER INTEGRALS

The concepts of (single and multiple) integral introduced earlier are not suitable for the unbounded domain of integration or for the unbounded integrand.

In this chapter we shall show how to extend the notion of integral to these two cases.

3.1. IMPROPER INTEGRALS OF THE FIRST KIND (ONE-DIMENSIONAL CASE)

In this section we shall extend the concept of definite integral to the one-dimensional *unbounded* connected domain of integration.

3.1.1. The improper integral of the first kind. One-dimensional unbounded connected domains are the half-lines $a \leqslant x < +\infty$, $-\infty < x \leqslant b$ and the infinite straight line $-\infty < x < +\infty$. Consider for definiteness $a \leqslant x < +\infty$.

Throughout this chapter, without specifying it in what follows, we shall assume that the function $f(x)$ is defined on the half-line $a \leqslant x < +\infty$ and that for any $R \geqslant a$ there is a definite integral $\int_a^R f(x)\,dx$ which we denote by the symbol $F(R)$:

$$F(R) = \int_a^R f(x)\,dx. \tag{3.1}$$

Thus under our hypotheses on the half-line $a \leqslant R < +\infty$ a function $F(R)$ defined by relation (3.1) is given. We look at the limiting value of $F(R)$ as $R \to +\infty$, i.e. examine the existence of the limit

$$\lim_{R \to +\infty} \int_a^R f(x)\,dx. \tag{3.2}$$

For the expression (3.2) we shall use the symbol

$$\int_a^\infty f(x)\,dx. \tag{3.3}$$

In what follows (3.3) will be called *improper integral* of the first kind of *a function* $f(x)$ over the half-line $a \leqslant x < +\infty$.

If the limit (3.2) exists, then the improper integral (3.3) is said to be *convergent*. If, however, the limit does not exist, the improper integral (3.3) is said to be *divergent*.

Remark 1. Consider the improper integral (3.3). If $b > a$, then along with this integral we may consider the integral $\int\limits_b^\infty f(x)\,dx$. Obviously, the convergence of one of these integrals implies the convergence of the other. The following equation holds:

$$\int\limits_a^\infty f(x)\,dx = \int\limits_a^b f(x)\,dx + \int\limits_b^\infty f(x)\,dx.$$

Note that the divergence of one of the improper integrals implies the divergence of the other.

Remark 2. If the improper integral (3.3) converges, the value of the limit (3.2) is denoted by the same symbol (3.3). Thus, if the integral (3.3) converges, we use the equation

$$\int\limits_a^\infty f(x)\,dx = \lim_{R \to +\infty} \int\limits_a^R f(x)\,dx.$$

Remark 3. The definition of the improper integrals $\int\limits_{-\infty}^b f(x)\,dx$ and $\int\limits_{-\infty}^{+\infty} f(x)\,dx$ is similar to that of the improper integral (3.3). The first of them symbolizes the operation of proceeding to the limit $\lim\limits_{R \to -\infty} \int\limits_R^b f(x)\,dx$, and the second symbolizes $\lim\limits_{\substack{R' \to -\infty \\ R'' \to +\infty}} \int\limits_{R'}^{R''} f(x)\,dx$.

Example. Consider on the half-line $a \leqslant x < \infty$ $(a > 0)$ the function $f(x) = 1/x^p$, $p = \text{const}$. It is integrable on any interval $a \leqslant x \leqslant R$, with

$$\int\limits_a^R \frac{dx}{x^p} = \begin{cases} \left.\dfrac{x^{1-p}}{1-p}\right|_a^R = \dfrac{R^{1-p} - a^{1-p}}{1-p} & \text{when } p \neq 1 \\ \ln x \Big|_a^R = \ln \dfrac{R}{a} & \text{when } p = 1. \end{cases}$$

Obviously, when $p > 1$ the limit $\lim\limits_{R\to\infty}\int\limits_a^R \frac{dx}{x^p}$ exists and is equal to $\frac{a^{1-p}}{p-1}$, and when $p \leqslant 1$ it does not exist. Consequently, the improper integral $\int\limits_a^\infty \frac{dx}{x^p}$ converges when $p > 1$ and diverges when $p \leqslant 1$. Note that when $p > 1$

$$\int\limits_a^\infty \frac{dx}{x^p} = \frac{a^{1-p}}{p-1}.$$

3.1.2. The Cauchy criterion for convergence of the improper integral of the first kind. Sufficient tests for convergence. The question of the convergence of the improper integral of the first kind is equivalent to the question of the existence of the limiting value of the function $F(R) = \int\limits_a^R f(x)\,dx$ as $R \to +\infty$. As is known*, for the limiting value of $F(R)$ as $R \to \infty$ to exist, it is necessary and sufficient that $F(R)$ should satisfy the following *Cauchy condition*: given any $\varepsilon > 0$ we can find $A > 0$ such that for any R' and R'' greater than A

$$|F(R'') - F(R')| = \left|\int\limits_{R'}^{R''} f(x)\,dx\right| < \varepsilon.$$

The following *statement* is thus true.

Theorem 3.1 (Cauchy criterion for convergence of the improper integral of the first kind). *For an improper integral* (3.3) *to converge it is necessary and sufficient that given any* $\varepsilon > 0$ *we should find* $A > 0$ *such that for any* R' *and* R'' *greater than* A

$$\left|\int\limits_{R'}^{R''} f(x)\,dx\right| < \varepsilon.$$

Remark. Note that the convergence of an improper integral does not imply even the boundedness of the integrand. For instance, the integral $\int\limits_0^\infty f(x)\,dx$, where the function is zero for nonintegral x and equals n for $x = n$ (an integer), clearly converges although the integrand is not bounded.

Since the Cauchy criterion is inconvenient for practical applications, it is appropriate to show the various sufficient tests for the convergence of improper integrals.

* See Section 8.1 of [1].

In what follows we shall assume that the function $f(x)$ is on $a \leqslant x < \infty$ and that for any $R \geqslant a$ there exists an ordinary integral $\int_a^R f(x)\,dx$.

We prove the following theorem.

Theorem 3.2 (general comparison test). Let on the half-line $a \leqslant x < \infty$

$$|f(x)| \leqslant g(x). \tag{3.4}$$

Then the convergence of $\int_a^\infty g(x)\,dx$ implies the convergence of $\int_a^\infty f(x)\,dx$.

Proof. Let $\int_a^\infty g(x)\,dx$ converge. Then, by the Cauchy criterion (see Theorem 3.1), given any $\varepsilon > 0$ we can find $A > 0$ such that for any $R' > A$ and $R'' > A$

$$\left|\int_{R'}^{R''} g(x)\,dx\right| < \varepsilon. \tag{3.5}$$

According to the ordinary inequalities for integrals and to inequality (3.4) we have

$$\left|\int_{R'}^{R''} f(x)\,dx\right| \leqslant \int_{R'}^{R''} |f(x)|\,dx \leqslant \int_{R'}^{R''} g(x)\,dx.$$

From this and from (3.5) it follows that for any R' and R'' greater than A

$$\left|\int_{R'}^{R''} f(x)\,dx\right| < \varepsilon.$$

Consequently, $\int_a^\infty f(x)\,dx$ converges

Theorem 3.3 (particular comparison test). Let on the half-line $0 < a \leqslant x < \infty$ the function $f(x)$ satisfy the relation

$$|f(x)| \leqslant \frac{c}{x^p},$$

where c and p are constants, $p > 1$. *Then* $\int_a^\infty f(x)\,dx$ *converges. If, however, there is a constant* $c > 0$ *such that on* $0 < a \leqslant x < \infty$ *we have* $f(x) \geqslant \frac{c}{x^p}$, *where* $p \leqslant 1$, *then* $\int_a^\infty f(x)\,dx$ *diverges.*

The statement of this theorem follows from Theorem 3.2 and the example of the preceding subsection (it suffices to put $g(x) = c/x^p$).

Corollary (particular comparison test in the limiting form). *If when* $p > 1$ *there is a finite limiting value* $\lim\limits_{x\to+\infty} |f(x)|\, x^p = c$, *then* $\int_a^\infty f(x)\,dx$ *converges. But if when* $p \leqslant 1$ *there is a positive limiting value* $\lim\limits_{x\to+\infty} f(x)\, x^p = c > 0$, *then* $\int_a^\infty f(x)\,dx$ *diverges.*

We shall show the validity of the first part of the corollary. To do this notice that the existence of a limit as $x \to +\infty$ implies the boundedness of the function $x^p\, |f(x)|$, i.e. given some constant $c_0 > 0$ we have

$$|f(x)| \leqslant c_0/x^p.$$

After that we apply the first part of Theorem 3.3. The validity of the second part of the corollary follows from the following arguments. Since $c > 0$, we can find $\varepsilon > 0$ so small that $c - \varepsilon > 0$. Corresponding to this ε is $A > 0$ such that when $x \geqslant A$ we have $c - \varepsilon < f(x)\, x^p$ (this inequality follows from the definition of the limit). Therefore $f(x) > \frac{c-\varepsilon}{x^p}$ and in this case the second part of Theorem 3.3 comes into play.

3.1.3. Absolute and conditional convergence of improper integrals. We introduce the concepts of *absolute* and *conditional* convergence of improper integrals. Let $f(x)$ be integrable over any closed interval $[a, R]$*.

Definition 1. *An improper integral* $\int_a^\infty f(x)\,dx$ *is said to be* **absolutely convergent** *if the integral* $\int_a^\infty |f(x)|\,dx$ *converges.*

---

* Then so is $|f(x)|$.

Definition 2. An improper integral $\int_a^\infty f(x)\,dx$ *is said to be conditionally convergent if it converges and* $\int_a^\infty |f(x)|\,dx$ *diverges.*

Remark. On putting $g(x) = |f(x)|$ in Theorem 3.2 we see that the absolute convergence of an improper integral implies its convergence.

Note that Theorems 3.2 and 3.3 establish only absolute convergence of the improper integrals under study.

Here is another sufficient test for convergence of improper integrals, suitable in the case of conditional convergence.

Theorem 3.4 (Abel-Dirichlet test). Let $f(x)$ *and* $g(x)$ *be functions defined on* $a \leqslant x < \infty$. *Also let* $f(x)$ *be continuous on* $a \leqslant x < \infty$ *and have on it a bounded antiderivative* $F(x)$*.

Suppose further that $g(x)$ *not increasing monotonically on* $a \leqslant x < \infty$ *tends to zero as* $x \to +\infty$ *and has a derivative* $g'(x)$ *continuous on* $a \leqslant x < \infty$. *Under these conditions the improper integral*

$$\int_a^\infty f(x)\,g(x)\,dx \tag{3.6}$$

converges.

Proof. We use the Cauchy criterion for convergence of improper integrals. As a preliminary we integrate by parts $\int_{R'}^{R''} f(x)\,g(x)\,dx$ on an arbitrary interval $[R', R'']$, $R'' > R'$, of the half-line $a \leqslant x < \infty$. We get

$$\int_{R'}^{R''} f(x)\,g(x)\,dx = F(x)\,g(x)\Big|_{R'}^{R''} - \int_{R'}^{R''} F(x)\,g'(x)\,dx. \tag{3.7}$$

Under the hypotheses of the theorem $F(x)$ is bounded: $|F(x)| \leqslant K$. Since $g(x)$ is not increasing and tends to zero as $x \to +\infty$, we have $g(x) \geqslant 0$ and $g'(x) \leqslant 0$. Thus, evaluating relation (3.7) we obtain the following inequality:

$$\left|\int_{R'}^{R''} f(x)\,g(x)\,dx\right| \leqslant K\,[g(R') + g(R'')] + K\int_{R'}^{R''} (-g'(x))\,dx.$$

---

* This means that $F(x)$, which may be defined to be $\int_a^x f(t)\,dt$, satisfies for all $x \geqslant a$ the inequality $|F(x)| \leqslant K$, where K is a constant.

Since the integral on the right of this is $g(R') - g(R'')$, obviously

$$\left|\int_{R'}^{R''} f(x)\,g(x)\,dx\right| \leqslant 2Kg(R'). \tag{3.8}$$

Using this inequality it is not hard to complete the proof of the theorem. Let ε be a positive number. Since $g(x) \to 0$ as $x \to +\infty$, for this ε we can choose A so that when $R' \geqslant A$ we have $g(R') < \varepsilon/2K$. From this and from (3.8) it follows that for any R' and R'' greater than A

$$\left|\int_{R'}^{R''} f(x)\,g(x)\,dx\right| < \varepsilon$$

which, by the Cauchy criterion, guarantees the convergence of (3.6). The proof of the theorem is complete.

Remark. The requirement on the differentiability of $g(x)$ in Theorem 3.4 is superfluous. Theorem 3.4 can be proved by assuming only that $g(x)$ is monotone and tends to zero as $x \to +\infty$; to do this we should use the second mean value formula (Bonnet formula).

Example 1. Consider the integral

$$\int_1^\infty \frac{\sin x}{x^\alpha}\,dx, \quad \alpha > 0. \tag{3.9}$$

Setting $f(x) = \sin x$, $g(x) = 1/x^\alpha$, it is easy to see that the hypotheses of Theorem 3.4 all hold for this integral. Therefore the integral (3.9) converges.

Example 2. Consider the Fresnel* integral $\int_0^\infty \sin x^2\,dx$. According to Remark 1 in Section 2.1.1 convergence of one of the integrals $\int_0^\infty \sin x^2\,dx$ and $\int_1^\infty \sin x^2\,dx$ implies the convergence of the other. We therefore turn to the second of the integrals. We have

$$\int_1^\infty \sin x^2\,dx = \int_1^\infty x \sin x^2 \frac{1}{x}\,dx.$$

Setting $f(x) = x \sin x^2$ and $g(x) = 1/x$ we can easily see that the hypotheses of Theorem 3.4 all hold and therefore the Fresnel integral converges.

---

* Augustin Fresnel (1788-1827) is an outstanding French physicist.

3.1.4. Change of variables under the improper integral and the formula for integration by parts. Now we shall formulate the conditions under which the formulas for change of variables and for integration by parts for improper integrals of the first kind are valid. Consider first the question of change of variable under the improper integral.

We shall assume the following conditions to hold:

(1) $f(x)$ *is continuous on the semiaxis* $a \leqslant x < \infty$;

(2) $a \leqslant x < \infty$ *is a set of values of some strictly monotone function* $x = g(t)$ *given on* $\alpha \leqslant t < \infty$ *(or* $-\infty < t \leqslant \alpha$*) and having on it a continuous derivative*;

(3) $g(\alpha) = a$.

Under these conditions convergence of one of the following improper integrals

$$\int_a^{\infty} f(x)\,dx \text{ and } \int_{\alpha}^{\infty} f(g(t))\,g'(t)\,dt \left(\text{or} - \int_{-\infty}^{\alpha} f(g(t))\,g'(t)\,dt\right) \qquad (3.10)$$

implies the convergence of the other and the equality of the integrals.

This statement is established using the following arguments.

Consider a closed interval $[a, R]$. Corresponding to it by the strict monotonicity of $g(t)$ is an interval $[\alpha, \rho]$ (or $[\rho, \alpha]$) of the axis t such that changing t on $[\alpha, \rho]$ results in the values of $x = g(t)$ filling $[a, R]$, with $g(\rho) = R$. Thus for those intervals all the conditions of Section 10.7.3 of [1] hold under which the formula for change of variable under the definite integral is valid. Therefore

$$\int_a^{R} f(x)\,dx = \int_{\alpha}^{\rho} f(g(t))\,g'(t)\,dt \left(\text{or} = -\int_{\rho}^{\alpha} f(g(t))\,g'(t)\,dt\right). \qquad (3.11)$$

By the strict monotonicity of $x = g(t)$, $R \to \infty$ as $\rho \to \infty$, and conversely $\rho \to \infty$ as $R \to \infty$ (or $R \to \infty$ as $\rho \to -\infty$ and $\rho \to -\infty$ as $R \to \infty$). Therefore formula (3.11) implies the validity of the above statement.

Now we proceed to discuss the question of integration by parts of improper integrals of the first kind.

We prove the following statement.

Let functions $u(x)$ *and* $v(x)$ *have continuous derivatives on* $a \leqslant x < \infty$, *and, in addition, let there exist a limiting value*

$$\lim_{x \to \infty} u(x)\,v(x) = A.$$

Under these conditions convergence of one of the integrals

$$\int_a^{\infty} u(x)\,v'(x)\,dx \text{ or } \int_a^{\infty} v(x)\,u'(x)\,dx \qquad (3.12)$$

implies the convergence of the other. Also

$$\int_a^\infty u(x)\,v'(x)\,dx = A - u(a)\,v(a) - \int_a^\infty v(x)\,u'(x)\,dx. \tag{3.13}$$

To prove the statement consider a closed interval $[a, R]$. The usual formula for integration by parts is valid on it. Therefore

$$\int_a^R u(x)\,v'(x)\,dx = [u(x)\,v(x)]_a^R - \int_a^R v(x)\,u'(x)\,dx.$$

Since as $R \to \infty$ the expression $[u(x)\,v(x)]_a^R$ tends to $A - u(a)\,v(a)$, the last equation implies simultaneous convergence or divergence of the integrals (3.12) and the validity of formula (3.13) in the case where one of the integrals (3.12) converges.

3.2. IMPROPER INTEGRALS OF THE SECOND KIND (ONE-DIMENSIONAL CASE)

In this section we extend the concept of definite integral to the case of unbounded functions.

3.2.1. The improper integral of the second kind. The Cauchy criterion. Let $f(x)$ be a function defined on a half-open interval $[a, b)$. The point b is said to be *singular* if $f(x)$ is not bounded on $[a, b)$ but is bounded on any closed interval $[a, b - \alpha]$ contained in $[a, b)$. It is also assumed that $f(x)$ is integrable on any such interval.

Under our hypotheses, on $(0, b - a]$ the function of the argument α is given defined by the relation

$$F(\alpha) = \int_a^{b-\alpha} f(x)\,dx.$$

We investigate the question of the right-hand limiting value of $F(\alpha)$ at the point $\alpha = 0$, i.e. question of the existence of a limit

$$\lim_{\alpha \to +0} \int_a^{b-\alpha} f(x)\,dx. \tag{3.14}$$

For the expression (3.14) we shall use the symbol

$$\int_a^b f(x)\,dx. \tag{3.15}$$

In what follows (3.15) will be called *an improper integral of the second kind* of a function $f(x)$ on a half-open interval $[a, b)$. If there exists a limit (3.14), then the improper integral (3.15) is said

to be *convergent*. But if that limit does not exist, then the improper integral (3.15) is said to be *divergent*. If the improper integral (3.15) converges, the value of the limit (3.14) is denoted by the same symbol (3.15). Thus, in case (3.15) converges we use the equation

$$\int_a^b f(x)\,dx = \lim_{\alpha\to 0} \int_a^{b-\alpha} f(x)\,dx.$$

Remark. The notion of improper integral of the second kind is easy to extend to the case where $f(x)$ has a finite number of singular points.

Example. Consider on $[a, b)$ a function $1/(b-x)^p$, $p > 0$. It is clear that b is a singular point for that function. It is also obvious that the function is integrable on any closed interval $[a, b-\alpha]$, with

$$\int_a^{b-\alpha} \frac{dx}{(b-x)^p} = \begin{cases} -\dfrac{(b-x)^{1-p}}{1-p}\Bigg|_a^{b-\alpha} = \dfrac{(b-a)^{1-p}-\alpha^{1-p}}{1-p} & \text{when } p \neq 1, \\ -\ln(b-x)\Bigg|_a^{b-\alpha} = \ln\dfrac{b-a}{\alpha} & \text{when } p = 1. \end{cases}$$

Clearly the limit $\lim\limits_{\alpha\to 0} \displaystyle\int_a^{b-\alpha} \frac{dx}{(b-x)^p}$ exists and is equal to $\dfrac{(b-a)^{1-p}}{1-p}$ when $p < 1$ and does not exist when $p \geqslant 1$. Consequently, the improper integral in question converges when $p < 1$ and diverges when $p \geqslant 1$.

We now state the Cauchy criterion for convergence of the improper integral of the second kind. We shall assume that $f(x)$ is given on $[a, b)$ and that b is a singular point of the function.

Theorem 3.5 (Cauchy criterion). For the improper integral of the second kind (3.15) to converge it is necessary and sufficient that given any $\varepsilon > 0$ we should be able to find $\delta > 0$ such that for any α' and α'' satisfying the condition $0 < \alpha'' < \alpha' < \delta$

$$\left| \int_{b-\alpha'}^{b-\alpha''} f(x)\,dx \right| < \varepsilon.$$

The validity of the theorem follows from the fact that the notion of convergence of an integral is by definition equivalent to the notion of the existence of a limiting value of $F(\alpha)$ introduced at the beginning of this subsection.

3.2.2. Concluding remarks. We shall not expound the theory of improper integrals of the second kind. This is because the main conclusions and theorems of the preceding section may be carried over without difficulty to the case of integrals of the second kind. Therefore we shall confine ourselves to some remarks.

1°. Under some constraints on integrands integrals of the second kind reduce to integrals of the first kind. Namely, let $f(x)$ be continuous on $[a, b)$ and let b be a singular point of the function. Under these conditions we can make the following change of variables in $\int_a^{b-\alpha} f(x)\,dx$:

$$x = b - \frac{1}{t}, \quad dx = \frac{dt}{t^2}, \quad \frac{1}{b-a} \leqslant t \leqslant \frac{1}{\alpha}.$$

As a result we get

$$\int_a^{b-\alpha} f(x)\,dx = \int_{1/(b-a)}^{1/\alpha} f\left(b - \frac{1}{t}\right)\frac{1}{t^2}\,dt. \tag{3.16}$$

Let $\int_a^b f(x)\,dx$ converge. This means that there exists a limit $\lim\limits_{\alpha\to+0} \int_a^{b-\alpha} f(x)\,dx$. Turning to equation (3.16) we see that there is also a limit as $1/\alpha \to +\infty$ of the expression on the right of (3.16). This proves that the improper integral of the first kind

$$\int_{1/(b-a)}^{\infty} f\left(b - \frac{1}{t}\right)\frac{1}{t^2}\,dt$$

converges and is equal to $\int_a^b f(x)\,dx$. Convergence of the above improper integral of the first kind clearly implies convergence of $\int_a^b f(x)\,dx$ and equality of the two integrals. So *convergence of one of the integrals*

$$\int_a^b f(x)\,dx \text{ or } \int_{1/(b-a)}^{\infty} f\left(b - \frac{1}{t}\right)\frac{1}{t^2}\,dt$$

implies convergence of the other and equality of the integrals.

2°. For improper integrals of the second kind it is easy to prove statements similar to those of Section 3.1.2 which may be combined under the general heading "comparison tests". Note that in all formulations $f(x)$ is considered on $[a, b)$, where b is a singular point of the function.

A particular comparison test will have the following form.

If $|f(x)| \leqslant c\,(b-x)^{-p}$, where $p < 1$, then the improper integral (3.15) converges. If, however, $f(x) \geqslant c\,(b-x)^{-p}$, where $c > 0$ and

$p \geqslant 1$, then the improper integral (3.15) diverges. The proof follows from the general comparison test and the example considered in the preceding subsection.

In analogy with Section 3.1.3 we could formulate for improper integrals rules of integration by change of variable and of integration by parts.

3.3. THE PRINCIPAL VALUE OF AN IMPROPER INTEGRAL

Definition. Let $f(x)$ be a function defined on a straight line $-\infty < x < \infty$ and integrable on every segment of that line. We shall say that $f(x)$ is integrable in the sense of Cauchy (Cauchy integrable) *if there is*

$$\lim_{R\to+\infty}\int_{-R}^{R} f(x)\,dx.$$

We shall call this limit the principal value *of the improper integral of $f(x)$ in the sense of Cauchy and denote it by the symbol**

$$\text{V.p.}\int_{-\infty}^{\infty} f(x)\,dx=\lim_{R\to+\infty}\int_{-R}^{R} f(x)\,dx.$$

Example 1. Find the principal value of the integral of $\sin x$. Since by the oddness of $\sin x$

$$\int_{-R}^{R}\sin x\,dx=0, \text{ we have V.p.}\int_{-\infty}^{\infty}\sin x\,dx=0.$$

The following *statement* is true.

If $f(x)$ is odd, then it is Cauchy integrable and the principal value of the integral of it is zero.

If $f(x)$ is even, then it is Cauchy integrable if and only if the improper integral

$$\int_{0}^{\infty} f(x)\,dx \tag{3.17}$$

converges.

The first part of the statement is obvious. To prove the second part it is sufficient to use the equation

$$\int_{-R}^{R} f(x)\,dx=2\int_{0}^{R} f(x)\,dx,$$

* V.p. are the initial letters of the French words *Valeur principal*, "principal value".

valid for any even function, and the definition of convergence of the improper integral (3.17).

The concept of Cauchy integrability may be introduced for improper integrals of the second kind also when the singular point is an interior point of the interval over which we integrate.

Definition. Let $f(x)$ be defined on a closed interval $[a, b]$, except possibly a point c, $a < c < b$, and integrable on any closed interval not containing c. We shall say that $f(x)$ is Cauchy integrable if there is a limit

$$\lim_{\alpha\to+0}\left(\int_a^{c-\alpha} f(x)\,dx+\int_{c+\alpha}^{b} f(x)\,dx\right)=\text{V.p.}\int_a^b f(x)\,dx$$

called **the principal value** *of the integral in the sense of Cauchy.*

Example 2. The function $\frac{1}{x-c}$ is not integrable on $[a, b]$, $a<c<b$ in the improper sense, it is Cauchy integrable, however. We have

$$\text{V.p.}\int_a^b \frac{dx}{x-c}=\lim_{\alpha\to+0}\left(\int_a^{c-\alpha}\frac{dx}{x-c}+\int_{c+\alpha}^{b}\frac{dx}{x-c}\right)=\ln\frac{b-c}{c-a}.$$

3.4. MULTIPLE IMPROPER INTEGRALS

This section extends the concept of multiple integral to the cases of the unbounded domain of integration and of the unbounded integrand. Recall that it is these cases that were omitted by us from consideration when constructing the theory of multiple integrals.

Note that we shall formulate the notion of improper multiple integral so that it embraces both the case of the unbounded domain of integration and the case of the unbounded function.

3.4.1. Multiple improper integrals. Let D be an open set* of an m-dimensional Euclidean space E^m. We shall use $\bar{D}$ to denote the closure of D obtained by adjoining to D its boundary. We shall need the concept of a sequence $\{D_n\}$ of open sets monotonically exhausting the set D.

We shall say that $\{\bar{D}_n\}$ monotonically exhausts D if: (1) *for any n the set $\bar{D}_n$ is contained in D_{n+1};* (2) *the union of all sets D_n coincides with D**.*

Note that each set D_n of $\{D_n\}$ is contained in D.

* A set is said to be open if it consists of interior points only. An open set is also called a domain.

** A union of all sets D_n is a set $\tilde{D}$ containing all the points of each of the sets D_n and such that each point of $\tilde{D}$ is in at least one of the sets D_n.

Suppose on D a function $f(x)$, $x = (x_1, x_2, \ldots, x_m)$ is given. Riemann integrable on any closed cubable subset of D. We shall consider all possible sequences $\{D_n\}$ of open sets monotonically exhausting D and having the property that the closure $\overline{D}_n$ of every set D_n is cubable (from which in particular it follows that each of the sets D_n is bounded).

If for any such sequence $\{D_n\}$ there exists a limit of the number sequence $\left\{\int\limits_{\overline{D}_n} f(x)\,dx\right\}$ and that limit is independent of the choice of sequence $\{D_n\}$, then the limit is said to be the improper integral of the function $f(x)$ over D and is denoted by one of the following symbols:

$$\int\limits_{D} f(x)\,dx, \quad \int\int\limits_{D}\cdots\int f(x_1, x_2, \ldots, x_m)\,dx_1\,dx_2 \ldots dx_m. \qquad (3.18)$$

The improper integral (3.18) *is called convergent.*

Note that the symbol (3.18) *is used also when the limits of sequences D_n* do not exist. Then the integral (3.18) is called *divergent*.

3.4.2. Improper integrals of nonnegative functions. We shall prove the following theorem.

Theorem 3.6. For the improper integral (3.18) *of a function $f(x)$ nonnegative in D to converge it is necessary and sufficient that for at least one sequence of cubable domains $\{D_n\}$ monotonically exhausting D the number sequence*

$$a_n = \int\limits_{\overline{D}_n} f(x)\,dx \qquad (3.19)$$

should be bounded.

Proof. The necessity of the hypothesis of the theorem is obvious: the sequence (3.19) is nondecreasing ($\overline{D}_n$ is contained in $\overline{D}_{n+1}$ and $f(x) \geqslant 0$), and therefore a necessary condition of its convergence is its boundedness. We proceed to prove the sufficiency of the hypotheses of the theorem. Since (3.19) is bounded and is not decreasing, it converges to some number I. It remains to prove that to the same number I converges the sequence

$$a'_n = \int\limits_{\overline{D}'_n} f(x)\,dx,$$

where $\{D'_n\}$ is another arbitrary sequence of domains monotonically exhausting D. Take any integer n and consider a domain D'_n. We

can find n_1 such that $\bar{D}'_n$ is contained in D_{n_1}*. Therefore

$$a'_n \leqslant a_{n_1} \leqslant I.$$

It follows that $\{a'_n\}$ converges to some number $I' \leqslant I$. Interchanging in our arguments the sequences a'_n and a_n we arrive at the inequality $I \leqslant I'$. Consequently, $I' = I$. This completes the proof of the theorem.

Example. Consider the integral

$$I = \iint_D e^{-x^2-y^2}\, dx\, dy \tag{3.20}$$

taken over an entire plane. Take the following system of concentric circles D_n to be a system of domains $\{D_n\}$ monotonically exhausting D:

$$x^2 + y^2 < n^2, \quad n = 1, 2, \ldots$$

In each of such circles D_n transform to a polar coordinate system r, φ. We get

$$a_n = \iint_{D_n} e^{-x^2-y^2}\, dx\, dy = \int_0^{2\pi} \int_0^n e^{-r^2} r\, dr\, d\varphi = \pi (1 - e^{-n^2}).$$

It follows that $\lim\limits_{n\to\infty} a_n = \pi$. By the theorem just proved the integral (3.20) converges and is equal to π. Note that the integral (3.20) may be represented as**

$$I = \int_{-\infty}^{\infty} e^{-x^2}\, dx \int_{-\infty}^{\infty} e^{-y^2}\, dy = \left(\int_{-\infty}^{\infty} e^{-x^2}\, dx \right)^2.$$

From this representation we obtain the value of the integral called a Poisson integral***:

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}.$$

* Suppose this is not the case. Then for any integer k we can find such a point M_k of the domain $\bar{D}'_n$ which is not in a domain D_k. We can choose (by virtue of the closure and boundedness of $\bar{D}'_n$) a subsequence of the sequence $\{M_k\}$ converging to some point $M \in \bar{D}'_n$. The point M, together with some neighbourhood, belongs to one of the sets D_{k_1}. But then belonging to the same set D_{k_1} and to all the sets D_k with large integers are points M_k with arbitrary large integers. But this contradicts the choice of point M_k.

** It is easy to see that such a representation is possible, if we take as the exhaustive system of domains a system of increasing squares with centres at the origin and with sides parallel to the axes and then apply the formula for iterated integration over each of such squares.

*** S.D. Poisson (1781-1840) is a French mathematician and physicist.

we may write

$$a_n = \omega_m \int_{1/n}^{a} r^{-p+m-1}\, dr.$$

It follows that the sequence a_n is bounded and hence converges if and only if $p < m$. By Theorem 3.6 the improper integral of $|x|^{-p}$ in D converges when $p < m$ and diverges when $p \geqslant m$.

Example 2. Let $a > 0$, D be the exterior of a ball of radius a with centre at the origin, and $g(x) = |x|^{-p}$. Take a system of concentric fibres D_n consisting of all the points $x \in E^m$ satisfying the condition

$$a < |x| < n$$

to be a sequence $\{D_n\}$ of domains monotonically exhausting D.

Using a spherical coordinate system we get

$$a_n = \int_{D_n} g(x)\, dx = \omega_m \int_{a}^{n} r^{-p+m-1}\, dr.$$

From this and Theorem 3.6 it follows that the improper integral of $|x|^{-p}$ in D converges when $p > m$ and diverges when $p \leqslant m$.

3.4.3. Improper integrals of functions that do not maintain sign. Here we shall discuss the relation between convergence and absolute convergence of multiple improper integrals. As in the one-dimensional case, the improper integral $\int_D f(x)\, dx$ is said to be absolutely convergent if the integral $\int_D |f(x)|\, dx$ converges. We shall prove that absolute convergence of an integral implies the usual convergence. The most remarkable is another property of multiple improper integrals that has no analogue in the one-dimensional case, which is that convergence of an improper multiple integral implies its absolute convergence. In other words, we shall prove that *for improper multiple integrals the concepts of convergence and of absolute convergence are equivalent.*

Before turning to the proof of these properties we shall make some preliminary remarks.

It follows from the definition of the improper integral that if the improper integral over D of each of the functions $f_+(x)$ and $f_-(x)$ converges, then so do the integrals of the sum or the difference of these functions.

Consider the following two nonnegative functions:

$$f_+(x) = \frac{|f(x)| + f(x)}{2}, \quad f_-(x) = \frac{|f(x)| - f(x)}{2}. \tag{3.21}$$

Obviously they may be defined by the relations

$$\left.\begin{aligned} f_+(x) &= \begin{cases} f(x), & \text{if } f(x) \geqslant 0, \\ 0, & \text{if } f(x) < 0, \end{cases} \\ f_-(x) &= \begin{cases} -f(x), & \text{if } f(x) \leqslant 0, \\ 0, & \text{if } f(x) > 0. \end{cases} \end{aligned}\right\} \tag{3.22}$$

Note also that the following obvious relations hold according to the definition of functions $f_+(x)$ and $f_-(x)$:

$$0 \leqslant f_+(x) \leqslant |f(x)|, \quad 0 \leqslant f_-(x) \leqslant |f(x)|, \tag{3.23}$$

$$f(x) = f_+(x) - f_-(x). \tag{3.24}$$

We now proceed to prove the statements made earlier in this subsection.

Theorem 3.8. The absolute convergence of a multiple improper integral $\int\limits_D f(x)\,dx$ *implies its usual convergence.*

Proof. We turn to the functions $f_+(x)$ and $f_-(x)$ just introduced. The integrability in the proper sense of $f(x)$ over any cubable subdomain of D implies the integrability over D of $|f(x)|$ and therefore from formulas (3.21) too it follows that $f_+(x)$ and $f_-(x)$ are also integrable over any such subdomain. Using the convergence of $\int\limits_D |f(x)|\,dx$, the just indicated property of $f_+(x)$ and $f_-(x)$, inequalities (3.23) and Theorem 3.7, it is easy to see that the improper integrals $\int\limits_D f_+(x)\,dx$ and $\int\limits_D f_-(x)\,dx$ converge. From this and from relation (3.24) it follows that so does $\int\limits_D f(x)\,dx$. The proof is complete.

Now we shall prove the converse.

Theorem 3.9. If a multiple improper integral $\int\limits_D f(x)\,dx$ *converges, it does so absolutely.*

Proof. Suppose the statement of the theorem is false. From Theorem 3.6 then it follows that the sequence of integrals of $|f(x)|$ over any sequence of domains $\{D_n\}$ monotonically exhausting D will be an infinitely large monotonically increasing sequence. It follows that $\{D_n\}$ may be chosen so that for any $n = 1, 2, \ldots$

$$\int\limits_{\bar{D}_{n+1}} |f(x)|\,dx > 3 \int\limits_{\bar{D}_n} |f(x)|\,dx + 2n. \tag{3.25}$$

Denote by P_n a set $D_{n+1} - D_n$. Then from (3.25) for any n we get

$$\int\limits_{\bar{P}_n} |f(x)|\,dx > 2 \int\limits_{\bar{D}_n} |f(x)|\,dx + 2n. \tag{3.26}$$

Since $|f(x)| = f_+(x) + f_-(x)$, we have

$$\int\limits_{\bar{P}_n} |f(x)|\,dx = \int\limits_{\bar{P}_n} f_+(x)\,dx + \int\limits_{\bar{P}_n} f_-(x)\,dx. \tag{3.27}$$

Suppose of the two integrals on the right of (3.27) the largest is the first. Then from relations (3.26) and (3.27), for any n

$$\int\limits_{\bar{P}_n} f_+(x)\,dx > \int\limits_{\bar{D}_n} |f(x)|\,dx + n. \tag{3.28}$$

Divide the domain P_n into a finite number of domains P_{n_i} so that the lower sum $\sum\limits_i m_i \Delta\sigma_i$ of $f_+(x)$ for this subdivision is different from the integral over $\bar{P}_n$ of $f_+(x)$ to so small an extent that changing on the left of (3.28) the integral of that lower sum yields the following inequality:

$$\sum_i m_i \Delta\sigma_i > \int\limits_{\bar{D}_n} |f(x)|\,dx + n. \tag{3.29}$$

Since $m_i \geqslant 0$, it is possible to leave in $\sum\limits_i m_i\Delta\sigma_i$ only the terms for which $m_i > 0$. The union of the corresponding domains P_{n_i} will be denoted by $\tilde{P}_n$.

In a domain $\tilde{P}_n$ the function $f(x)$ is positive and therefore $f(x) = f_+(x)$ in it. By (3.29) therefore we have

$$\int\limits_{\bar{\tilde{P}}_n} f(x)\,dx > \int\limits_{\bar{D}_n} |f(x)|\,dx + n. \tag{3.30}$$

Denote by D_n^* the union of D_n and $\tilde{P}_n$. Adding inequality (3.30) to the obvious inequality

$$\int\limits_{\bar{D}_n} f(x)\,dx > -\int\limits_{\bar{D}_n} |f(x)|\,dx,$$

we then have

$$\int\limits_{\bar{D}_n^*} f(x)\,dx > n. \tag{3.31}$$

Clearly the sequence of domains $\{D_n^*\}$ monotonically exhausts the domain D. But then, by (3.31), $\int_D f(x)\,dx$ diverges. Since under the hypothesis this integral converges, the assumption that the statement of the theorem is false does not hold. This completes the proof of the theorem.

3.4.4. The principal value of multiple improper integrals.

Definition. Let $f(x)$ be defined for all $x \in E^m$ and integrable in every ball K_R of radius R with centre at the origin. We shall say that $f(x)$ is **Cauchy integrable** *in E^m if there exists a limit*

$$\lim_{R\to\infty} \int_{K_R} f(x)\,dx.$$

This limit is called the **principal value** *of the improper integral of $f(x)$ in the sense of Cauchy and designated*

$$\text{V.p.} \int_{E^m} f(x)\,dx = \lim_{R\to\infty} \int_{K_R} f(x)\,dx.$$

Example. Let $f(x)$ have in spherical coordinates the form $f(x) = h(r)\,g(\theta_1, \theta_2, \ldots, \theta_{m-1})$, where the functions h and g are continuous, with

$$\int_0^{2\pi} d\theta_1 \int_0^{\pi} d\theta_2 \ldots \int_0^{\pi} g(\theta_1, \theta_2, \ldots, \theta_{m-1}) \left(\prod_{k=1}^{m-1} \sin^{k-1}\theta_k \right) d\theta_{m-1} = 0.$$

Then clearly $f(x)$ is Cauchy integrable and

$$\text{V.p.} \int_{E^m} f(x)\,dx = 0.$$

In particular, for $m = 2$ the function of two variables $f(x, y) = h(r)\cos\varphi$ is Cauchy integrable and the integral of it in the sense of the principal value is zero.

In the case where $f(x)$ has a singularity at some point x_0 of D the integral in the sense of Cauchy is defined as the limit

$$\text{V.p.} \int_D f(x)\,dx = \lim_{R\to\infty} \int_{D_R} f(x)\,dx,$$

where D_R is a set obtained by removing from D a ball of radius R with centre at x_0.

CHAPTER 4

LINE INTEGRALS

In this chapter we extend the concept of one-dimensional definite integral taken over a line segment to the case where the domain of integration is a segment of some plane or space curve.

Integrals of this kind are called *line integrals*. In applications it is customary to consider line integrals of two kinds (of expressions having a scalar and a vector meaning). In this chapter line integrals of the first and the second kind are discussed simultaneously.

4.1. DEFINITION OF LINE INTEGRALS AND THEIR PHYSICAL INTERPRETATION

Consider in the Oxy plane some rectifiable curve L having neither points of self-intersection nor overlappings. Suppose that the curve is defined by the parametric equations

$$\begin{cases} x=\varphi(t), \\ y=\psi(t) \end{cases} \quad (a \leqslant t \leqslant b) \tag{4.1}$$

and first consider it to be unclosed, with end points A and B.

Suppose further that

the function $f(x, y)$	the two functions $P(x, y)$ and $Q(x, y)$

are defined and continuous along the curve $L = AB$*.

Divide the closed interval $a \leqslant t \leqslant b$ using the points $a = t_0 < t_1 < t_2 < \ldots < t_n = b$ into n subintervals $[t_{k-1}, t_k]$ $(k = 1, 2, \ldots, n)$.

* The function $f(x, y)$ is said to be *continuous along* L if given any $\varepsilon > 0$ we can find $\delta > 0$ such that $|f(x_1, y_1) - f(x_2, y_2)| < \varepsilon$ for any two points (x_1, y_1) and (x_2, y_2) of L satisfying the condition $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} < \delta$. In fact we have defined not the continuity but the uniform continuity of $f(x, y)$ along L, but since the set of all points of L is bounded and closed these notions coincide.

8—01684

Since on L a certain point M_k (x_k, y_k) with coordinates $x_k = \varphi(t_k)$, $y_k = \psi(t_k)$ corresponds to each value of t_k, this subdivision of $a \leqslant t \leqslant b$ divides the curve $L = AB$ into n subarcs M_0M_1, M_1M_2, . . ., $M_{n-1}M_n$ (Fig. 4.1).

Choose on every subarc $M_{k-1}M_k$ an arbitrary point N_k (ξ_k, η_k) whose coordinates ξ_k, η_k correspond to some value τ_k of the parameter t, so that $\xi_k = \varphi(\tau_k)$, $\eta_k = \psi(\tau_k)$, with $t_{k-1} \leqslant \tau_k \leqslant t_k$. Let us agree to denote by Δl_k the length of the kth subarc $M_{k-1}M_k$ $(k = 1, 2, \ldots, n)$.

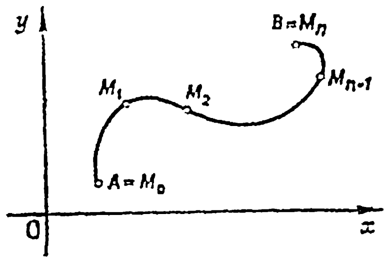


Fig. 4.1

Form the integral sum

$$\sigma_1 = \sum_{k=1}^{n} f(\xi_k, \eta_k)\cdot\Delta l_k. \tag{4.2}$$

Form the two integral sums

$$\sigma_2 = \sum_{k=1}^{n} P(\xi_k, \eta_k)(x_k - x_{k-1}), \tag{4.2'}$$

$$\sigma_3 = \sum_{k=1}^{n} Q(\xi_k, \eta_k)(y_k - y_{k-1}). \tag{4.2''}$$

The number I is said to be *the limit* of the integral sum σ_s $(s = 1, 2, 3)$ as the largest of the lengths Δl_k tends to zero, if given any $\varepsilon > 0$ we can find $\delta > 0$ such that $|\sigma_s - I| < \varepsilon$ as soon as the largest of the lengths Δl_k is less than δ.

Definitions

If there exists a limit of the integral sum σ_1 as the largest of the lengths Δl_k tends to zero, then this limit is said to be a line integral of the first kind *of $f(x, y)$ along L and designated*

$$\int_L f(x, y)\,dl$$

If there exists a limit of the integral sum σ_2 $[\sigma_3]$ as the largest of the lengths Δl_k tends to zero, then this limit is said to be a line integral of the second kind *and designated*

$$\int_{AB} P(x, y)\,dx \left[\int_{AB} Q(x, y)\,dy\right].$$

or

$$\int_{AB} f(x, y)\, dl. \qquad (4.3)$$

It is customary to call the sum

$$\int_{AB} P(x, y)\, dx + \int_{AB} Q(x, y)\, dy$$

the general line integral of the second kind and to designate it by the symbol

$$\int_{AB} P(x, y)\, dx + Q(x, y)\, dy. \qquad (4.3')$$

Now we shall give a *physical interpretation* of the line integrals we have introduced.

Suppose a mass with linear density $f(x, y)$ is distributed along the curve L. To compute the mass of the entire curve, it is natural to divide the curve into small segments and, assuming the density to vary little on each segment, to set the mass of each segment approximately equal to the product of some intermediate value of density by the length of that segment.

In such a case the mass of the entire curve will be approximately equal to the integral sum (4.2). The exact value of the mass is naturally defined to be the limit of the sum (4.2) as the length of the largest segment tends to zero.

Thus *the line integral of the first kind* (4.3) *gives the mass of a curve the linear density along which is equal to* $f(x, y)$.

Suppose a particle moves along the curve L from A to B under the force $\vec{F}(x, y)$ with components $P(x, y)$ and $Q(x, y)$. To compute the work done by the force to move the particle, it is natural to divide the curve L into small segments and, assuming the force to vary little on each segment, to set the work on each segment approximately equal to the sum of the products of the force components taken at some intermediate points by the displacement vector components.

In such a case all the work done by the force on the particle to move it from A to B will be approximately equal to the sum of (4.2′) and (4.2″). The exact value of the work is naturally defined to be the limit of that sum as the length of the largest segment tends to zero.

Thus the *general line integral of the second kind* (4.3′) *gives the work done by a force, with components* $P(x, y)$ *and* $Q(x, y)$, *on a particle to move it along the curve* L *from* A *to* B.

Remark 1. It is obvious from the form of the sums (4.2), (4.2′) and (4.2″) that the line integral of the first kind is independent

of the direction in which (from A to B or from B to A) the curve L is travelled while reversing the sense of the curve results in reversing the sign, i.e.

$$\int\limits_{\breve{AB}} P(x, y)\,dx = -\int\limits_{\breve{BA}} P(x, y)\,dx, \quad \int\limits_{\breve{AB}} Q(x, y)\,dy = -\int\limits_{\breve{BA}} Q(x, y)\,dy.$$

Remark 2. Quite similarly introduced for the space curve are the line integral of the first kind $\int\limits_{\breve{AB}} f(x, y, z)\,dl$ and the three line integrals of the second kind

$$\int\limits_{\breve{AB}} P(x, y, z)\,dx, \quad \int\limits_{\breve{AB}} Q(x, y, z)\,dy, \quad \int\limits_{\breve{AB}} R(x, y, z)\,dz.$$

It is customary to call the sum of the last three integrals the *general line integral of the second kind* and to designate it by the symbol

$$\int\limits_{\breve{AB}} P(x, y, z)\,dx + Q(x, y, z)\,dy + R(x, y, z)\,dz.$$

4.2. EXISTENCE OF LINE INTEGRALS AND REDUCING THEM TO DEFINITE INTEGRALS

Let us agree to say that the curve L is *smooth* if the functions $\varphi(t)$ and $\psi(t)$ of the parametric equations (4.1) defining it have on $[a, b]$ continuous derivatives $\varphi'(t)$ and $\psi'(t)$*.

The curve L will be said to be *piecewise smooth* if it is continuous and splits into a finite number of pieces with no interior points in common, each piece being a smooth curve.

According to what was agreed upon as far back as Chapters 15 and 16 of [1], the points corresponding to the value of the parameter t for which the derivatives $\varphi'(t)$ and $\psi'(t)$ both vanish will be called singular points of the curve L.

We shall prove that *if the curve $L = AB$ is smooth and does not contain singular points and if the functions $f(x, y)$, $P(x, y)$ and $Q(x, y)$ are continuous along the curve, then the following formulas*

---

* By this we mean that the derivatives $\varphi'(t)$ and $\psi'(t)$ are continuous at any interior point of $[a, b]$ and have finite limiting values at the point a from the right and at b from the left.

reducing line integrals to usual definite integrals are valid

$$\int\limits_{AB} f(x, y)\, dl = \int\limits_a^b f[\varphi(t), \psi(t)] \times \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt. \quad (4.4)$$

$$\int\limits_{AB} P(x, y)\, dx = \int\limits_a^b P[\varphi(t), \psi(t)]\, \varphi'(t)\, dt, \quad (4.4')$$

$$\int\limits_{AB} Q(x, y)\, dy = \int\limits_a^b Q[\varphi(t), \psi(t)]\, \psi'(t)\, dt. \quad (4.4'')$$

At the same time we shall prove the existence of all line integrals occurring in these formulas.

We notice first that the definite integrals on the right of (4.4), (4.4′), and (4.4″) a fortiori exist (for under our hypotheses the integrands in each of these integrals are continuous on $a \leqslant t \leqslant b$).

For the line integral of the second kind we shall derive only formula (4.4′) (for derivation of (4.4″) is quite similar).

As in Section 4.1 we divide the interval $a \leqslant t \leqslant b$ by means of the points $a = t_0 < t_1 < t_2 < \ldots < t_n = b$ into n subintervals and form the integral sums (4.2) and (4.2′).

Now we take into account the fact that

$$\Delta l_k = \int\limits_{t_{k-1}}^{t_k} \sqrt{[\varphi'(t)]^2 + [\varphi'(t)]^2}\, dt.$$

$$x_k - x_{k-1} = \varphi(t_k) - \varphi(t_{k-1}) = \int\limits_{t_{k-1}}^{t_k} \varphi'(t)\, dt.$$

This allows us to rewrite the expressions for (4.2) and (4.2′) as follows

$$\sigma_1 = \sum_{k=1}^{n} \left\{ f[\varphi(\tau_k), \psi(\tau_k)] \times \int\limits_{t_{k-1}}^{t_k} \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt \right\}. \quad (4.5)$$

$$\sigma_2 = \sum_{k=1}^{n} \left\{ P[\varphi(\tau_k), \psi(\tau_k)] \times \int\limits_{t_{k-1}}^{t_k} \varphi'(t)\, dt \right\}. \quad (4.5')$$

(We have also taken into account the fact that $\xi_k = \varphi(\tau_k)$, $\eta_k = \psi(\tau_k)$, where τ_k is some value of t satisfying the condition $t_{k-1} \leqslant \tau_k \leqslant t_k$.)

Now denote the definite integrals on the right-hand sides of formulas (4.4) and (4.4') by K_1 and K_2, respectively. Dividing $a \leqslant t \leqslant b$ into a sum of n subintervals $[t_{k-1}, t_k]$ we may write the definite integrals K_1 and K_2 as follows:

$$K_1 = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f[\varphi(t), \psi(t)] \times \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt.$$

$$K_2 = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} P[\varphi(t), \psi(t)] \times \varphi'(t)\, dt.$$

Consider and evaluate the differences

$$\sigma_1 - K_1 = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \{f[\varphi(\tau_k), \psi(\tau_k)] - f[\varphi(t), \psi(t)]\} \times \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt. \quad (4.6)$$

$$\sigma_2 - K_2 = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \{P[\varphi(\tau_k), \psi(\tau_k)] - P[\varphi(t), \psi(t)]\}\, \varphi'(t)\, dt. \quad (4.6')$$

Since $x = \varphi(t)$ and $y = \psi(t)$ are continuous on $a \leqslant t \leqslant b$ and $f(x, y)$ and $P(x, y)$ are continuous along L, by the theorem on the continuity of a composite function (see Section 14.3 of [1]) $f[\varphi(t), \psi(t)]$ *and* $P[\varphi(t), \psi(t)]$ *are continuous on* $a \leqslant t \leqslant b$.

Now notice that as the largest of the lengths of the subarcs Δl_k tends to zero, so does the largest of the differences $(t_k - t_{k-1})$*. But from this it follows that given any $\varepsilon > 0$ we can find $\delta > 0$ such that provided the largest of the lengths Δl_k is less than δ, each of the braces in formulas (4.6) and (4.6') is less than ε. Therefore, provided the largest of the lengths Δl_k is less than δ we obtain for

* Indeed $\Delta l_k = \int_{t_{k-1}}^{t_k} \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt$. Since $\varphi'(t)$ and $\psi'(t)$ are continuous on $a \leqslant t \leqslant b$ and do not vanish together, the function $\sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}$ is continuous and strictly positive on $a \leqslant t \leqslant b$. Therefore the minimum value m of the last function on $a \leqslant t \leqslant b$ is positive. But then $\Delta l_k > m \int_{t_{k-1}}^{t_k} dt = m(t_k - t_{k-1})$, i.e. $t_k - t_{k-1} \leqslant \frac{1}{m} \Delta l_k$.

the differences (4.6) and (4.6′) the following evaluations:

$$|\sigma_1 - K_1| \leqslant \varepsilon \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \times$$
$$\times \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt =$$
$$= \varepsilon \int_a^b \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt =$$
$$= \varepsilon l,$$

where l is the length of the curve AB.

$$|\sigma_2 - K_2| \leqslant$$
$$\leqslant \varepsilon \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} |\varphi'(t)|\, dt \leqslant$$
$$\leqslant \varepsilon M \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} dt = \varepsilon M (b-a),$$

where M is the maximum value of $|\varphi'(t)|$ on $a \leqslant t \leqslant b$. We stress that in deriving (4.4′) we require only that $\varphi'(t)$ should be continuous and that the curve $L = AB$ should be rectifiable (it is not required that $\psi'(t)$ should be continuous).

In view of the arbitrariness of ε we may assert that the integral sums σ_1 and σ_2 have (as the largest of the lengths Δl_k tends to zero) limits equal to K_1 and K_2 respectively. This simultaneously proves the existence of line integrals on the left-hand sides of (4.4) and (4.4′) and the validity of these formulas.

Remark 1. *In the case of the piecewise smooth curve L it is* natural to define line integrals along the curve as the sums of the corresponding line integrals over all smooth pieces forming the curve L. Equations (4.4), (4.4′) and (4.4″) thus turn out to be true for the piecewise smooth curve L too. They are also true in the case where $f(x, y)$, $P(x, y)$ and $Q(x, y)$ are not strictly continuous but merely piecewise continuous along L (i.e. where L splits into a finite number of pieces, having no interior points in common, along each of which the functions are continuous).

Remark 2. Quite similar results and formulas are true also for line integrals taken over the space curve $L = AB$ defined by the parametric equations

$$\begin{cases} x = \varphi(t), \\ y = \psi(t), \quad (a \leqslant t \leqslant b). \\ z = \chi(t), \end{cases}$$

We confine ourselves to writing the formulas

$$\int_{AB} f(x, y, z)\,dl = \int_a^b f[\varphi(t), \psi(t), \chi(t)] \times \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2 + [\chi'(t)]^2}\,dt.$$

$$\int_{AB} P(x, y, z)\,dx = \int_a^b P[\varphi(t), \psi(t), \chi(t)]\,\varphi'(t)\,dt,$$

$$\int_{AB} Q(x, y, z)\,dy = \int_a^b Q[\varphi(t), \psi(t), \chi(t)]\,\psi'(t)\,dt,$$

$$\int_{AB} R(x, y, z)\,dz = \int_a^b R[\varphi(t), \psi(t), \chi(t)]\,\chi'(t)\,dt.$$

Remark 3. We have established above that the line integral of the second kind depends on the sense of rotation of the curve $L = AB$. We should therefore specify what we mean by the symbol

$$\int_L P(x, y)\,dx + Q(x, y)\,dy \qquad (4.7)$$

when L is a *closed* curve (i.e. when the point B coincides with A).

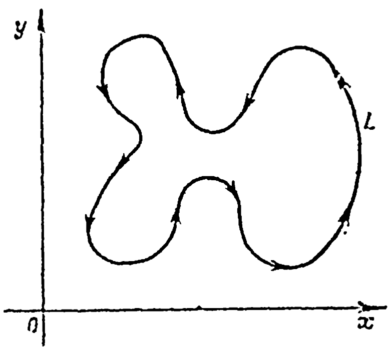


Fig. 4.2

Of the two possible directions of circulation about a closed contour L the direction in which the domain inside the contour remains to the left of the point tracing around the path is called *positive**. In Fig. 4.2 the positive direction is represented by arrows.

We shall assume that in the integral (4.7) along the closed contour L this contour is always traced in the positive direction.

Remark 4. It is easy to show that *line integrals have the same property that the usual definite integrals have* (the proofs are similar to those presented in Sections 10.5 and 10.6 of [1]). Under more stringent hypotheses, however, these properties follow at once from formulas (4.4), (4.4′) and (4.4″).

* By convention such a direction of motion may be called a "counterclockwise motion".

We list them as applied to line integrals of the first kind.

1°. Linear property. If for either of the functions $f(x, y)$ and $g(x, y)$ there is a line integral along the curve AB and if α and β are any constants, then for the function $[\alpha f(x, y) + \beta g(x, y)]$ there is also a line integral along AB, with

$$\int_{AB} [\alpha f(x, y)+\beta g(x, y)]\, dl=\alpha \int_{AB} f(x, y)\, dl+\beta \int_{AB} g(x, y)\, dl.$$

2°. Additivity. If the arc AB is made up of two arcs AC and CB and if for $f(x, y)$ there is a line integral along AB, then for $f(x, y)$ there is a line integral along either of the arcs AC and CB, with

$$\int_{AB} f(x, y)\, dl= \int_{AC} f(x, y)\, dl+ \int_{CB} f(x, y)\, dl.$$

3°. The computation of the absolute value of the integral. If there is a line integral along AB of $f(x, y)$, then there is a line integral along AB of $|f(x, y)|$, with

$$\left|\int_{AB} f(x, y)\, dl\right| \leqslant \int_{AB} |f(x, y)|\, dl.$$

4°. Mean value formula. If $f(x, y)$ is continuous along the curve AB, then there is a point M^* on that curve such that

$$\int_{AB} f(x, y)\, dl=l\cdot f(M^*),$$

where l is the length of AB.

Example 1°. Compute the mass of the ellipse L defined by the parametric equations

$$\begin{cases} x=a\cos t, \\ y=b\sin t \end{cases} \quad (0\leqslant t\leqslant 2\pi)$$

provided $a > b > 0$ and the linear density of mass distribution is $\rho = |y|$.

The problem reduces to computing the line integral of the first kind $\int_L |y|\, dl$.

Using formula (4.4) we get

$$\int_L |y|\, dl=b\int_0^{2\pi} |\sin t|\sqrt{a^2\sin^2 t+b^2\cos^2 t}\, dt=$$

$$=b\int_0^{\pi} \sin t\sqrt{a^2\sin^2 t+b^2\cos^2 t}\, dt-$$

$$-b\int_{\pi}^{2\pi}\sin t\sqrt{a^2\sin^2 t+b^2\cos^2 t}\,dt=$$

$$=-b\int_{0}^{\pi}\sqrt{a^2-(a^2-b^2)\cos^2 t}\,d(\cos t)+$$

$$+b\int_{\pi}^{2\pi}\sqrt{a^2-(a^2-b^2)\cos^2 t}\,d(\cos t)=2b\left(b+a\,\frac{\arcsin e}{e}\right),$$

where* $e=\frac{\sqrt{a^2-b^2}}{a}$.

Example 2°. Compute the line integral of the second kind

$$I=\int_L (x^2-2xy)\,dx+(y^2-2xy)\,dy$$

where L is the parabola $y=x^2$, with $-1\leqslant x\leqslant 1$. This parabola may be regarded as a curve given by the parametric equations

$$\begin{cases} x=t, \\ y=t^2 \end{cases} \quad (-1\leqslant t\leqslant 1).$$

Using formulas (4.4′) and (4.4″) therefore we have

$$I=\int_{-1}^{1}(t^2-2t^3)\,dt+\int_{-1}^{1}(t^4-2t^3)\,2t\,dt=-\frac{14}{15}.$$

* Recall that the number e is termed *eccentricity* in analytic geometry.

CHAPTER 5

SURFACE INTEGRALS

In this chapter we shall consider integration of functions given on surfaces. In this connection we first investigate the concept of surface and the concept of surface area.

5.1. THE SURFACE

5.1.1. The surface. The mapping f of a domain* G in the plane onto the set C^* of three-dimensional Euclidean space is said to be *homeomorphic* if f is a one-to-one correspondence between the points of G and G^* under which corresponding to any convergent sequence $\{M_n\}$ of points of G is a convergent sequence $\{M_n^*\}$ of points of G^* and corresponding to every convergent sequence $\{M_n^*\}$ of points of G^* is a convergent sequence of points $\{M_n\}$ of G. In other words, a homeomorphic mapping of the domain G onto the set G^* is a one-to-one continuous mapping of these sets. We shall say that G^* is the *image* of G under homeomorphic mapping f.

Consider the following example. Let G be a domain in the Oxy plane, (u, v) the coordinates of a point M of that domain, $z = z(M)$ a function continuous in G, and G^* the graph of that function. Clearly the mapping f of G onto G^* given by the relations

$$x = u, \; y = v, \; z = z(u, v)$$

is a homeomorphic mapping of that domain onto G^*.

We introduce the concept of *elementary surface.*

A set Φ of points of a three-dimensional space is said to be an elementary surface *if it is the image of an open disk G under homeomorphic mapping of G into the space***.

Using the notion of elementary surface we introduce the concept of the so-called *simple surface.*

Before proceeding we introduce the notion of neighbourhood of a point of a set Φ of an Euclidean space E^3.

---

* Recall that a *domain* is a set each point of which is an interior point.

** We are considering three-dimensional Euclidean space, although we could consider Euclidean space of any number of dimensions and speak of two-dimensional surface in that space.

A neighbourhood of a point M of Φ is the common part of Φ and the spatial neighbourhood of the point M.

A set Φ of points of a space is said to be a **simple surface** *if that set is connected* and any point of the set has a neighbourhood that is an elementary surface.*

Note that an elementary surface is a simple surface but that a simple surface is not an elementary surface in general. For instance, the sphere is a simple but nonelementary surface.

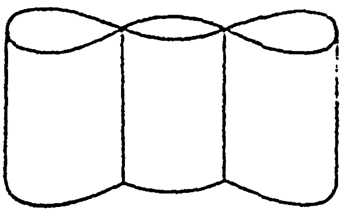


Fig. 5.1

We formulate the concept of *general surface.*

The mapping f of a simple surface G is said to be *locally homeomorphic* if each point of G has a neighbourhood that is homeomorphically mapped onto its image.

A set Φ of points of a space is said to be a **general surface** *if it is the image of a simple surface under its locally homeomorphic mapping into the space.*

Remark 1. Note that neighbourhoods of points on a general surface are introduced as the images of neighbourhoods of points of that simple surface whose image is a given general surface.

Remark 2. A simple surface is clearly a surface without self-intersections and without overlappings. A general surface may have self-intersections and overlappings. For instance, the surface in Fig. 5.1 has self-intersections but is the locally homeomorphic image of a zone of a cylinder and is therefore a general surface.

5.1.2. The regular surface. We introduce the concept of regular (k-times differentiable) surface.

A surface Φ whose points have coordinates x, y, z is said to be **regular** (k-**times differentiable**) *if for some $k \geqslant 1$ each point of Φ has a neighbourhood allowing a k-time differentiable parametrization.* This means that each of the above neighbourhoods is a homeomorphic mapping of some elementary domain G** into the plane (u, v) by means of the relations

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v) \tag{5.1}$$

in which the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ are k-times differentiable in G.

If $k = 1$, then the surface is usually said to be *smooth.*

* Recall that a set is said to be connected if any two of its points can be joined by a continuous curve consisting entirely of points of that set.

** A domain G in the plane is said to be *elementary* if it is the image of an open disk under homeomorphic mapping of that disk onto the plane.

We shall also say that by means of relations (5.1) we introduce in a neighbourhood of a point on the surface *a regular parametrization* using the parameters u and v.

Remark 1. If the whole of Φ is given by the mapping (5.1) of G, then we shall say that *a single parametrization* is introduced on Φ.

A point of a regular surface is said to be *ordinary* if there is a regular parametrization of some neighbourhood of it such that at that point the rank of the matrix

$$A = \begin{pmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{pmatrix} \tag{5.2}$$

is two. Otherwise the point of the surface is called *a singular point.*

A domain G in the plane is said to be *simple* if that domain is a simple plane surface. For instance, the open ring is a simple domain.

We shall say that the function $f(u, v)$ belongs to the class C^k in G if it is k-times differentiable and if all its partial derivatives of order k are continuous in G.

The following theorem is true.

Theorem 5.1. Let G be a simple domain in the plane (u, v) and let $x(u, v)$, $y(u, v)$, $z(u, v)$ be functions of class C^k, $k \geqslant 1$, given in G, with the rank of the matrix (5.2) equalling two at all points of G. Then relations (5.1) define in space a set Φ which is a regular k-times differentiable general surface without singular points.

Proof. Obviously it suffices to show that using relations (5.1) gives a locally homeomorphic mapping of G onto Φ.

Let $M_0(x_0, y_0, z_0)$ be any fixed point of Φ corresponding to the values (u_0, v_0) of the parameters (u, v) (Fig. 5.2). Under the hypothesis the rank of the matrix A is two at a point (u_0, v_0). Suppose for definiteness the determinant $\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}$ of the matrix A is nonzero at that point. Since the determinant is a Jacobian $\frac{D(x, y)}{D(u, v)}$ and is nonzero at the point (u_0, v_0) and $x(u, v)$, $y(u, v)$ have continuous partial derivatives in G, by the theorem on the solvability of a system of functional equations (see Theorem 15.2 in [1]) there is a neighbourhood H of the point (x_0, y_0) in the Oxy plane such that in that neighbourhood there is a unique and k-times differentiable solution

$$u = u(x, y), \quad v = v(x, y) \tag{5.3}$$

of the system

$$\left.\begin{aligned} x(u, v) - x = 0, \\ y(u, v) - y = 0. \end{aligned}\right\}$$

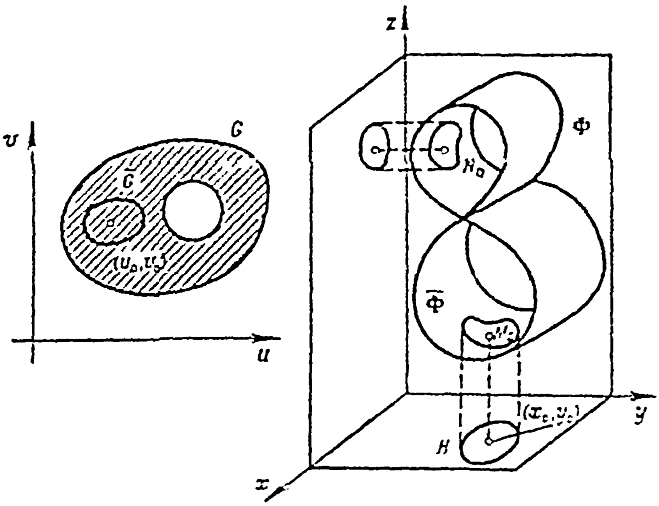


Fig. 5.2

It follows from the above reasoning that some neighbourhood H of the point (x_0, y_0) is a homeomorphic mapping of some neighbourhood $\bar{G}$ of the point (u_0, v_0) with the aid of the relations $x = x(u, v)$, $y = y(u, v)$ (the inverse mapping of H onto $\bar{G}$ is effected using relations (5.3)).

Substituting expressions (5.3) for u and v in the relation $z = z(u, v)$ we see that some neighbourhood $\bar{\Phi}$ of a point M_0 on Φ is the graph of the k-times differentiable function $z = z(u(x, y), v(x, y)) = z(x, y)$. But this means that using the function $z(x, y)$ effects a homeomorphic mapping of a neighbourhood H of a point (x_0, y_0) in the Oxy plane onto the neighbourhood $\bar{\Phi}$ of the point M_0 of Φ. It is obvious that the neighbourhood $\bar{G}$ of the point (u_0, v_0) is homeomorphically mapped onto the neighbourhood $\bar{\Phi}$ of M_0 on Φ*. In other words Φ is the image of G under locally homeomorphic mapping into space and is therefore a general surface. This completes the proof.

Remark 2. In the process of proving the theorem we have established that *each point M_0 of a surface Φ without singular points has a neighbourhood $\bar{\Phi}$ which is 1-1 projected onto one of the coordinate planes and is therefore the graph of a k-times differentiable function* (this was the function $z(x, y)$ in the proof of the theorem).

* Here we have used the obvious statement that a composition of homeomorphic mappings results in a homeomorphic mapping.

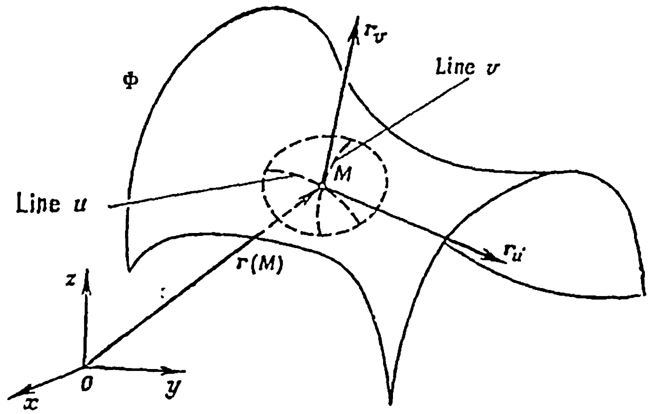


Fig. 5.3

Figure 5.2 shows points M_0 and N_0 whose neighbourhoods are 1-1 projected onto the Oxy and Oxz planes respectively.

5.1.3. Defining a surface by means of vector functions. Consider a regular surface Φ. It is a certain set of points M with coordinates (x, y, z) (Fig. 5.3). Denote by $\boldsymbol{r}(M)$ the vector going from the origin to a point M on the surface. Clearly $\boldsymbol{r}(M)$ is the *vector function* of a variable point M of the surface*. It is usually called the *radius vector* of the surface Φ.

Let us turn to that neighbourhood of the point M which is the image of a homeomorphic mapping (5.1) of some elementary domain G** (in Fig. 5.3 the neighbourhood is encircled by a broken line). Then clearly the coordinates $x(u, v)$, $y(u, v)$, $z(u, v)$ of the point M are the coordinates of the vector $\boldsymbol{r}(M)$. It is obvious that in this neighbourhood the function $\boldsymbol{r}(M)$ is the function of the variables u and v: $\boldsymbol{r}(M) = \boldsymbol{r}(u, v)$. With the value of v fixed, the terminal point of the radius vector $\boldsymbol{r}(u, v)$ traces in the neighbourhood under consideration a curve called a *curve* u (or a *curve* v = const). With the value of u fixed, the terminal point of $\boldsymbol{r}(u, v)$ traces a *curve* v (or u = = const). These curves u and v are called *coordinate curves* on the surface Φ in the neighbourhood under consideration.

A system of coordinate curves u and v may thus be introduced in some neighbourhood of each point of the surface Φ. It is sometimes

* A vector function may be regarded as a collection of three scalar functions. Detailed information on vector functions is given in Section 12.1. We shall use it as need arises.

** A domain G in the plane is said to be elementary if it is the homeomorphic image of an open disk.

called a *system of curvilinear coordinates on the surface* (more precisely, in the neighbourhood under consideration).

In Section 12.1 a geometrical interpretation of the derivatives r_u and r_v of the vector function $r(u, v)$ is given. These vectors are the tangent vectors of coordinate curves (see Fig. 5.3).

Using the vectors r_u and r_v one can see the geometrical meaning of the ordinary and the singular point of the regular surface.

Recall that a point M of a surface is said to be ordinary if in a neighbourhood of the point we can introduce a parametrization

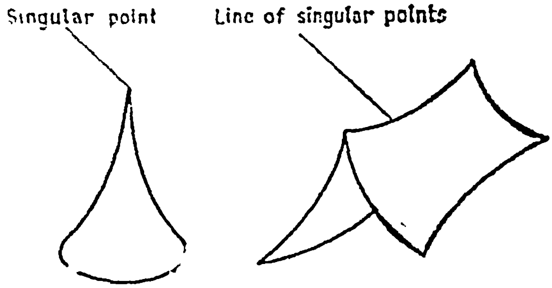


Fig. 5.4

with the aid of equations (5.1) such that the rank of the matrix A (see relation (5.2)) at the point is 2. Since the rows of A consist of the coordinates of r_u and r_v and the rank of A is two, r_u and r_v are linearly independent. So the ordinary point is characterized by the possibility of introducing in a neighbourhood of the point a parametrization such that r_u and r_v are linearly independent at the point M.

In Fig. 5.3 M is an ordinary point of the surface Φ. Figure 5.4 represents the surfaces with singular points.

5.1.4. The tangential plane and the normal to a surface. One-sided and two-sided surfaces. We have already introduced the concept of tangential plane to a surface which is the graph of the differentiable function $z = z(x, y)$ (see Section 14.4.2 of [1]). Recall that the tangential plane at a point M_0 was defined to be a plane having the property that the angle between that plane and the secant M_0M (M being an arbitrary point of the surface) tends to zero as M tends to M_0. We have proved that if $z(x, y)$ is a function differentiable at a point (x_0, y_0), then there is a tangential plane at the point $M_0(x_0, y_0, z(x_0, y_0))$.

We shall show that there is a tangential plane at any ordinary point of a smooth surface. To do this it is clearly enough to establish that some neighbourhood of the ordinary point of the surface is the graph of the differentiable function. But in Section 5.1.2 (see Remark there) we proved this property for any ordinary point of

a smooth surface. Consequently, *there is a tangential plane at any ordinary point of a smooth surface.*

Remark 1. It follows from the definition of the tangential plane to a surface Φ that the tangent line at a point M_0 to any smooth curve* situated on the surface and passing through M_0 lies in the tangential plane to Φ at M_0. Since r_u and r_v are tangential to the curves u and v passing through M_0, they are in the tangential plane at M_0.

We introduce the concept of *normal* to a surface Φ at a point M_0.

The *normal* to a surface Φ at a point M_0 is a straight line passing through M_0 and perpendicular to the tangential plane at M_0. The *normal vector* to the surface at M_0 is any nonzero vector collinear with the normal at M_0.

Let M_0 be an ordinary point of a smooth surface Φ and suppose that some neighbourhood $\overline{\Phi}$ of that point is defined using a vector function $r(u, v)$ such that the vectors r_u and r_v are not collinear at M_0. Then clearly the vector

$$N = [r_u r_v] \tag{5.4}$$

is the normal vector to Φ and the vector

$$n = \frac{[r_u r_v]}{|[r_u r_v]|} \tag{5.5}$$

is the unit normal vector to Φ.

Remark 2. Since under the hypothesis the surface is smooth, the vector function $N(u, v)$ and the vector function $n(u, v)$ defined by relations (5.4) and (5.5) respectively are continuous. *There is thus a continuous normal vector field at some neighbourhood of each point of a smooth surface.*

The question naturally arises: is there global continuous vector field of normals on any smooth surface? It turns out that there are surfaces on which there are no global continuous vector fields of normals. An example of such a surface is the so-called Möbius strip** depicted in Fig. 5.5. (This surface is obtained from the rectangle $ABB'A'$ by pasting together the sides AB and $A'B'$ in such a way that the points A and B' and the points A' and B coincide, see Fig. 5.5.)

A surface on which there is a global continuous vector field of normals is called *two-sided.* A surface on which there is no such global field is called *one-sided.*

The plane, sphere, ellipsoid, one-sheeted hyperboloid are two-sided surfaces, the Möbius strip being a one-sided surface.

---

* A curve L is said to be *smooth* if it can be given using the vector function $r(t)$ of class C^1 for which $r'(t) \neq 0$ (for more detail see Section 12.2).

** A. Möbius (1790-1868) is a German mathematician.

We shall consider only two-sided surfaces in what follows.

5.1.5. Auxiliary lemmas. Here we prove some statements we shall need in further discussion.

Lemma 1. Let M_0 be an ordinary point of a smooth surface Φ. Then some neighbourhood of M_0 is 1-1 projected onto the tangential plane drawn at any point of the neighbourhood.

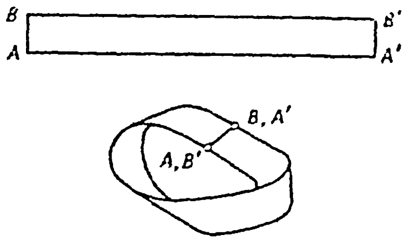


Fig. 5.5

Proof. We show that for instance that neighbourhood $\overline{\Phi}$ of M_0 possesses the property pointed out in the lemma in which the normal at any point makes with the normal at M_0 an angle less than $\pi/4$ and which is 1-1 projected onto some disk in one of the coordinate planes (for instance, in the Oxy plane)*. Note first that the normals at any two points of $\overline{\Phi}$ make an angle less than $\pi/2$. Further let $\overline{\Phi}$ lack the above property. Then for some point M of $\overline{\Phi}$ we can find points P and Q of $\overline{\Phi}$ such that the chord PQ is parallel to the normal n_M at M (Fig. 5.6). Consider the line of intersection of $\overline{\Phi}$ with a plane parallel to the Oz axis and passing through PQ. By the choice of the neighbourhood $\overline{\Phi}$ the portion PNQ of the line lies in $\overline{\Phi}$ and is the graph of the differentiable function given on the segment which is the projection of PQ onto the Oxy plane. By the Lagrange theorem the tangent at some point N of PNQ is parallel to the chord

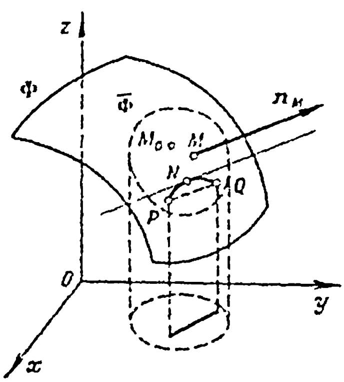


Fig. 5.6

* The possibility of choosing such a neighbourhood $\overline{\Phi}$ stems from the following considerations. In Section 5.1.4 we noted (see Remark 2) that there is a continuous vector field of normals at some neighbourhood of an ordinary point on a surface. In a sufficiently small neighbourhood of M_0 therefore the normals make with the normal at M_0 an angle less than $\pi/4$. We have also established that some neighbourhood of M_0 is 1-1 projected onto a coordinate plane. In that neighbourhood there is clearly a part that is projected onto some disk in the coordinate plane.

PQ and therefore to the normal n_M at M. But then the normal at N perpendicular to that tangent makes an angle of $\pi/2$ with the normal at M. But this is not possible since the normals at any two points of $\overline{\Phi}$ (including M and N) make an angle less than $\pi/2$. The contradiction obtained shows that the lemma is valid, which completes the proof of the lemma.

We introduce the concept of *complete* surface. *A surface* Φ *is said to be complete if any fundamental sequence of points of* Φ *converges to some point of the surface.*

The plane, sphere, ellipsoid, one-sheeted hyperboloid are examples of complete surfaces. The open disk, any open connected set on a sphere are incomplete surfaces. Bounded complete surfaces and bounded closed parts of complete surfaces will be called *bounded complete surfaces* in what follows.

We shall say that the part of Φ has a size less than δ if it can be placed in some sphere whose diameter is less than δ.

The following lemma holds.

Lemma 2. Let Φ *be a smooth bounded complete surface without singular points. There is* $\delta > 0$ *such that any part of* Φ *whose size is less than* δ *is 1-1 projected onto the tangential plane passing through any point of that part.*

Proof. Suppose the statement of the lemma is false. Then for any $\delta_n = 1/n$, $n = 1, 2, \ldots$, there is a part Φ_n of Φ whose size is less than δ_n and which is not 1-1 projected onto the tangential plane at some point of it. Choose in every part Φ_n a point M_n and select a subsequence of $\{M_n\}$ converging to some point M_0 of Φ*. Consider a neighbourhood of M_0 satisfying the conditions of Lemma 1. With n sufficiently large, the neighbourhood will contain all parts Φ_n. But then every part should be 1-1 projected onto the tangential plane (at any of its points) but this contradicts the choice of parts Φ_n. The proof of the lemma is complete.

The following lemma is true.

Lemma 3. Let Φ *be a smooth bounded complete surface without singular points. There is* $\delta > 0$ *such that any part of* Φ *whose size is less than* δ *is 1-1 projected onto one of the coordinate planes.*

The proof of the lemma is closely analogous to that of Lemma 2.

Lemma 4. Let Φ *be a smooth bounded complete two-sided surface without singular points. Then given any* $\varepsilon > 0$ *we can find* $\delta > 0$ *such that for the cosine of the angle* γ *between the unit normal vectors at any two points of an arbitrary part* $\overline{\Phi}$ *of the surface whose size is less than* δ *we have the representation*

$$\cos \gamma = 1 - \alpha_{\Phi}, \tag{5.6}$$

where $|\alpha_{\Phi}| < \varepsilon$.

* Since Φ is a bounded complete surface, such a subsequence can be chosen.

Proof. Consider a vector field of unit normals $n(M)$ continuous on Φ (such a field exists, since Φ is a two-sided surface). The vector function n is uniformly continuous since Φ is a bounded complete surface and hence a bounded closed set. Given any $\varepsilon > 0$ therefore we can find $\delta > 0$ such that for two arbitrary points M_1 and M_2 of Φ the distance between which is less than δ we have

$$|n(M_2) - n(M_1)| < \sqrt{2\varepsilon}. \tag{5.7}$$

Since

$$\cos\gamma = 1 - \frac{1}{2}(n(M_2) - n(M_1))^2\,*,$$

putting

$$\alpha_\Phi = \frac{1}{2}(n(M_2) - n(M_1))^2$$

and using inequality (5.7) we see that relations (5.6) are true. This completes the proof of the lemma.

5.2. SURFACE AREA

5.2.1. Surface area. Let Φ be a bounded complete two-sided surface. Divide Φ with piecewise smooth curves into a finite number of parts Φ_i each 1-1 projected onto the tangential plane passing through any point of that part**. Denote by Δ the maximum size of the parts Φ_i and by σ_i the area of the projection of Φ_i onto the tangential plane at some point M_i of the part Φ_i. Further form the sum $\sum_i \sigma_i$ of all the areas.

We state the following *definitions*.

Definition 1. The number σ is said to be the limit *of the sum $\sum_i \sigma_i$ as $\Delta \to 0$ if given any $\varepsilon > 0$ we can find $\delta > 0$ such that for all subdivisions of Φ with piecewise smooth curves into a finite number of parts Φ_i for which $\Delta < \delta$ regardless of the choice of points M_i on the parts Φ_i*

$$\left|\sum_i \sigma_i - \sigma\right| < \varepsilon. \tag{5.8}$$

---

* We have used the following relations:

$$n^2(M_1) = 1,\quad n^2(M_2) = 1,\quad n(M_2)\,n(M_1) = \cos\gamma,$$

$$\frac{1}{2}(n(M_2) - n(M_1))^2 = \frac{1}{2}(n^2(M_2) - 2n(M_2)\,n(M_1) + n^2(M_1)).$$

** The possibility of such a subdivision is guaranteed by Lemma 2 of Section 5.1.5.

Definition 2. If for a surface Φ there is a limit σ of the sums $\sum_i \sigma_i$ *as $\Delta \to 0$, then the surface is said to be* squarable *and the number σ is said to be the surface area.*

Our immediate task is to find sufficient conditions for the squarability of a surface. We shall prove that smooth bounded complete two-sided surfaces are squarable. We shall simultaneously show the computational techniques used to compute surface areas.

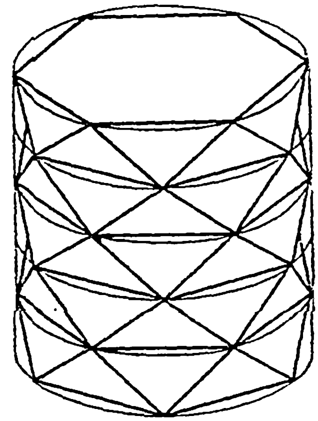


Fig. 5.7

At first sight it would be natural to approach the question of computing the area of a surface using approximation of the surface with polyhedra. This way is ineffectual, however. We shall show an example due to Schwarz* which demonstrates that the areas of polyhedra inscribed into a smooth surface may increase unboundedly as the number of faces increases and their size decreases.

Let Φ be a zone of a cylinder (Fig. 5.7). Divide Φ with circles parallel to the bases of Φ into n equal parts. Next divide each of such circles into m equal parts as shown in Fig. 5.7. Depicted in the figure is also a polyhedron Φ_{nm} inscribed into Φ. With any m fixed, the area of that polyhedron Φ_{nm} clearly exceeds by a factor of n that of the projection of the polyhedron onto the plane of the cylinder base. Since the projection does not depend on n, by increasing n, with any m fixed, the area of the polyhedron Φ_{nm} may be *made arbitrarily large.*

5.2.2. Squarability of smooth surfaces. We shall prove the following theorem.

Theorem 2. A smooth bounded complete two-sided surface without singular points is squarable.

Proof. Suppose a single regular parametrization may be introduced on a surface Φ. In that case the radius vector $r(M)$ of a variable point Φ of the surface is the function $r(u, v)$ of the class C^{1**} given in some closed bounded domain Ω in the plane of variables u and v. The partial derivatives r_u and r_v of $r(u, v)$ are continuous vector functions independent of the choice of Cartesian coordinate system in space. The value σ of $\iint_\Omega |[r_u, r_v]|\, du\, dv$ therefore is independent of the choice of Cartesian coordinate system in space. We prove that Φ is squarable and that its area is σ.

---

* H.A. Schwarz (1843-1921) is a German mathematician.

** By this it should be understood that each component of $r(u, v)$ belongs to C^1.

Let ε be an arbitrary positive number taken to be fixed in further reasoning. Find for that $\varepsilon > 0$ a number $\delta > 0$ proceeding from the following requirements: (1) any part Φ_i of Φ whose size is less than δ is projected uniquely onto the tangential plane at any point of the part Φ_i; (2) the cosine of the angle γ between the unit normal vectors at any two points of the part Φ_i may be represented by

$$\cos\gamma = 1 - \alpha_{\Phi_i}, \tag{5.9}$$

where $|\alpha_{\Phi_i}| < \varepsilon/\sigma$ and $|\alpha_{\Phi_i}| < 1$. The possibility of such a choice of $\delta > 0$ is guaranteed by Lemmas 2 and 4 of Section 5.1.5.

Consider an arbitrary subdivision of Φ by means of piecewise smooth curves into a finite number of parts Φ_i whose maximum size Δ is not greater than δ. Since there is a single parametrization on Φ, corresponding to this subdivision of Φ into parts Φ_i is subdivision of a domain Ω into parts Ω_i. On every part Φ_i choose an arbitrary point M_i and denote by σ_i the area of the projection of a part Φ_i onto the tangential plane at M_i. To compute σ_i proceed as follows. Choose a Cartesian coordinate system so that its origin coincides with M_i, the Oz axis is directed along the normal vector to the surface at M_i and the Ox and Oy axes are in the tangential plane. In our coordinate system the surface is defined by the parametric equations $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$, and the vector $[r_u, r_v]$ has the coordinates $\{A, B, C\}$, where

$$A = \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}, \quad B = \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \quad C = \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}. \tag{5.10}$$

Note that for the points of a part Φ_i, in view of the choice of δ and the orientation of the Oz axis, the number C is positive, $C > 0$. Also note that the cosine of the angle γ_M between the normal at a point M of Φ_i and the Oz axis is

$$\cos\gamma_M = \frac{C}{|[r_u r_v]|}. \tag{5.11}$$

It is clear that γ_M is the angle between the normals at the points M and M_i of the part Φ_i and therefore representation (5.9) is true for it.

We now turn to the integral $\iint\limits_{\Omega_i} |[r_u r_v]|\, du\, dv$ which is clearly independent of the choice of Cartesian coordinates in space. Using the positivity of C and the third of the formulas (5.10) we get

$$\iint\limits_{\Omega_i} |[r_u r_v]|\, du\, dv = \iint\limits_{\Omega_i} \frac{|[r_u r_v]|}{\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}} \left|\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}\right| du\, dv. \tag{5.12}$$

Applying to the integral on the right of (5.12) the first mean value formula in generalized form we have

$$\iint\limits_{\Omega_i} |[r_u r_v]|\, du\, dv = \left(\frac{|[r_u r_v]|}{\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}}\right)_M \iint\limits_{\Omega_i} \left|\frac{\mathscr{D}(x, y)}{\mathscr{D}(u, v)}\right| du dv, \tag{5.13}$$

where M is some point of Φ_i.

Since

$$\left(\frac{|[r_u r_v]|}{\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}}\right)_M = \frac{1}{\cos \gamma_M}$$

(see (5.10) and (5.11)), and $\iint\limits_{\Omega_i} \left|\frac{\mathscr{D}(x, y)}{\mathscr{D}(u, v)}\right| du\, dv = \sigma_i^*$, from formula (5.13) and representation (5.9) for $\cos \gamma_M$ we find that

$$\sigma_i = \iint\limits_{\Omega_i} |[r_u r_v]|\, du\, dv - \iint\limits_{\Omega_i} \alpha_{\Phi_i} |[r_u r_v]|\, du\, dv. \tag{5.14}$$

Adding up equations (5.14) for all parts Φ_i and considering that $\sum\limits_i \iint\limits_{\Omega_i} |[r_u r_v]|\, du\, dv = \iint\limits_{\Omega} |[r_u r_v]|\, du\, dv = \sigma$ we get

$$\sum_i \sigma_i = \sigma - \sum_i \iint\limits_{\Omega_i} \alpha_{\Phi_i} |[r_u r_v]|\, du\, dv. \tag{5.15}$$

Evaluate the last term on the right of (5.15). We have

$$\left|\sum_i \iint\limits_{\Omega_i} \alpha_{\Phi_i} |[r_u r_v]|\, du\, dv\right| \leqslant \sum_i \iint\limits_{\Omega_i} |\alpha_{\Phi_i}|\, |[r_u r_v]|\, du\, dv <$$

$$< \frac{\varepsilon}{\sigma} \sum_i \iint\limits_{\Omega_i} |[r_u r_v]|\, du\, dv = \frac{\varepsilon}{\sigma} \cdot \sigma = \varepsilon.$$

From this and from equation (5.15) we get

$$\left|\sum_i \sigma_i - \sigma\right| < \varepsilon.$$

Thus the surface Φ is squarable and its area equals σ.

We have considered the case where it is possible to introduce a single parametrization on Φ. In the general case Φ may be divided into a finite number of parts, each allowing a single parametrization

* We have used the formula for the area of a plane domain in transforming from coordinates (x, y) to coordinates (u, v) using the relations $x = x(u, v)$, $y = y(u, v)$.

to be introduced*. After that the area of the surface may be defined as the sum of the areas of those parts. This completes the proof of the theorem.

Remark 1. Let the surface Ф be piecewise smooth, i.e. made up of a finite number of smooth bounded complete two-sided surfaces. The surface Ф is clearly squarable, its area may be defined as the sum of the areas of the constituent surfaces.

Remark 2. In the process of proving Theorem 5.2 we have established that if it is possible to introduce a single parametrization on Ф and if the domain of the radius vector $r(u, v)$ of Ф is a closed bounded domain Ω in the plane (u, v), then the area σ of Ф can be found from the formula

$$\sigma = \iint_{\Omega} |[r_u r_v]| \, du \, dv. \tag{5.16}$$

If $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ are parametric equations of the surface, then the vector $[r_u r_v]$ has the coordinates $\{A, B, C\}$ defined by relations (5.10). Since $|[r_u r_v]| = \sqrt{A^2 + B^2 + C^2}$, formula (5.16) may be written as

$$\sigma = \iint_{\Omega} \sqrt{A^2+B^2+C^2} \, du \, dv. \tag{5.17}$$

Using the symbols

$$r_u^2 = E, \quad r_u r_v = F, \quad r_v^2 = G$$

and the formula

$$|[r_u r_v]| = \sqrt{r_u^2 r_v^2 - (r_u r_v)^2},$$

expression (5.16) for the area of a surface may also be written as

$$\sigma = \iint_{\Omega} \sqrt{EG - F^2} \, du \, dv. \tag{5.18}$$

Remark 3. The area of a surface has the additive property: *if a surface Ф is divided by a piecewise smooth curve into parts* $Ф_1$ *and* $Ф_2$ *having no interior points in common, then the area* σ *of* Ф *is equal to the sum* $\sigma_1 + \sigma_2$ *of the areas of* $Ф_1$ *and* $Ф_2$. This property follows from the representation of area using the integral and from the additive property of the integral.

* We may use for example Lemma 3 of Section 5.1.5. By this lemma Ф can be divided into a finite number of parts, each of which is 1-1 projected onto some coordinate plane and thus is the graph of a differentiable function.

5.3. SURFACE INTEGRALS

5.3.1. Surface integrals of the first and the second kind. Let Φ be a smooth bounded complete two-sided surface. Let a function $f(M)$ of a point M be given on Φ. Denote by $n(M)$ a continuous vector field of unit normals to Φ.

We divide Φ with piecewise smooth curves into parts Φ_i and choose on each of such parts an arbitrary point M_i. We introduce the following notation: Δ is the maximum size of the parts Φ_i, σ_i is the area of Φ_i, X_i, Y_i, Z_i are the angles a vector $n(M_i)$ makes with the coordinate axes.

Form the following four sums:

$$I\{\Phi_i, M_i\} = \sum_i f(M_i)\,\sigma_i, \tag{5.19}$$

$$I\{\Phi_i, M_i, Z_i\} = \sum_i f(M_i)\cos Z_i\sigma_i, \tag{5.20}$$

$$I\{\Phi_i, M_i, Y_i\} = \sum_i f(M_i)\cos Y_i\sigma_i, \tag{5.21}$$

$$I\{\Phi_i, M_i, X_i\} = \sum_i f(M_i)\cos X_i\sigma_i. \tag{5.22}$$

For each of these sums we introduce the concept of limit as $\Delta \to 0$. We formulate this concept for the sums (5.19). For the sums (5.20), (5.21), and (5.22) the notion of limit is formulated in a similar way.

Definition. The number I is said to be the limit of a sum $I\{\Phi_i, M_i\}$ *as $\Delta \to 0$ if given any $\varepsilon > 0$ we can find $\delta > 0$ such that for any subdivisions of the surface Φ with piecewise smooth curves into a finite number of parts Φ_i whose maximum size Δ is less than δ, regardless of the choice of points M_i on parts Φ_i we have*

$$|I\{\Phi_i, M_i\} - I| < \varepsilon.$$

The limit I of a sum $I\{\Phi_i, M_i\}$ as $\Delta \to 0$ is called *a surface integral of the first kind of the function $f(M)$* over Φ and designated

$$I = \iint_{\Phi} f(M)\,d\sigma. \tag{5.23}$$

If (x, y, z) are the coordinates of a point M on Φ, then for $f(M)$ we may use the symbol $f(x, y, z)$. In that case formula (5.23) may be written as

$$I = \iint_{\Phi} f(x, y, z)\,d\sigma. \tag{5.24}$$

The limits of the sums $I\{\Phi_i, M_i, Z_i\}$, $I\{\Phi_i, M_i, Y_i\}$, and $I\{\Phi_i, M_i, X_i\}$ as $\Delta \to 0$ are called *surface integrals of the second*

kind of the function $f(M)$ over Φ. For these integrals we use the following symbols respectively:

$$\iint_{\Phi} f(M)\cos Z\,d\sigma, \quad \iint_{\Phi} f(M)\cos Y\,d\sigma, \quad \iint_{\Phi} f(M)\cos X\,d\sigma$$

or symbols similar to (5.24).

Remark 1. The definition of the surface integral of the first kind implies *the independence of the integral from the choice of orientation of the vector field of unit normals to the surface* or, as we say, *from the choice of side of the surface.*

Remark 2. *The surface integral of the second kind depends on the choice of side of the surface*: reversing the orientation of the vector field of unit normals reverses the sign of all the three surface integrals of the second kind. This is due to the fact that in each of the sums (5.20), (5.21), and (5.22) the values of $f(M_i)$ and σ_i remain unchanged under a change of orientation, and the values of the cosines of the angles the normal $n(M_i)$ makes with the coordinate axes change sign.

Remark 3. Once a definite side of the surface is chosen, the surface integrals of the second kind may obviously be regarded as the surface integrals of the first kind over Φ of the functions $f(M)\cos Z(M)$, $f(M)\cos Y(M)$, $f(M)\cos X(M)$ respectively. Indeed, once a definite side of the surface is chosen, $\cos Z$, $\cos Y$, $\cos X$ are the functions of the point M of Φ.

5.3.2. Existence of surface integrals of the first and the second kind. Let Φ be a surface satisfying the conditions stated at the beginning of Section 5.3.1. Choose on Φ a definite side. By Remark 3 of Section 5.3.1, once a definite side of Φ is chosen surface integrals of the second kind may be regarded as integrals of the first kind. Therefore we shall state sufficient conditions for existence only for the integrals of the first kind.

The following theorem is true.

Theorem 5.3. Let it be possible to introduce on a surface Φ a single parametrization by the function

$$x = x(u, v), \quad y = y(u, v), \quad z = z(u, v) \tag{5.25}$$

given in a bounded closed domain Ω of the plane (u, v) and belonging to the class C^1 in that domain. If the function $f(M) = f(x, y, z)$ is continuous on Φ, then the surface integral of the first kind of that*

* The concept of continuity of the function of a point M given on some set $\{M\}$ in space was formulated in Section 14.3.1 of [1]. In the case under consideration the role of $\{M\}$ is played by Φ.

function over Φ exists and can be evaluated from the formula

$$I=\iint_{\Phi} f(M)\,d\sigma=$$

$$=\iint_{\Omega} f(x(u, v), y(u, v), z(u, v))\sqrt{EG-F^2}\,du\,dv^*. \qquad (5.26)$$

Proof. We want to prove that given any $\varepsilon>0$ we can find $\delta>0$ such that for any subdivision of Φ with piecewise smooth curves into a finite number of parts Φ_i for which $\Delta<\delta$, regardless of the choice of points M_i on the parts Φ_i we have

$$\Big| I\{\Phi_i, M_i\}-$$

$$-\iint_{\Omega} f(x(u, v), y(u, v), z(u, v))\sqrt{EG-F^2}\,du\,dv\Big|<\varepsilon. \qquad (5.27)$$

Let ε be any fixed positive number. Choose for this $\varepsilon>0$ a number $\delta^*>0$ so that the following two conditions should hold:

(1) For any two points $(\tilde{u}_i, \tilde{v}_i)$ and (u_i, v_i) of Ω a distance less than δ^* apart,

$$\Big|\sqrt{E(\tilde{u}_i, \tilde{v}_i)G(\tilde{u}_i, \tilde{v}_i)-F^2(\tilde{u}_i, \tilde{v}_i)}-$$

$$-\sqrt{E(u_i, v_i)G(u_i, v_i)-F^2(u_i, v_i)}\Big|<\frac{\varepsilon}{2AP}, \qquad (5.28)$$

where A is a positive number exceeding the maximum of the function $|f(M)|$ and P is the area of Ω;

(2) For any subdivision of Ω with piecewise curves into a finite number of parts Ω_i whose size is less than δ^* and for any choice of points (u_i, v_i) in every part Ω_i

$$\Big|\sum_i f(x(u_i, v_i), y(u_i, v_i), z(u_i, v_i))\times$$

$$\times\sqrt{E(u_i, v_i)G(u_i, v_i)-F^2(u_i, v_i)}\,\sigma_i^*-$$

$$-\iint_{\Omega} f(x(u, v), y(u, v), z(u, v)\sqrt{EG-F^2}\,du\,dv\Big|<\frac{\varepsilon}{2}, \qquad (5.29)$$

where σ_i^* are the areas of parts Ω_i.

The possibility of the required choice of δ^* is guaranteed by the property of uniform continuity of the function $\sqrt{EG-F^2}$ continuous

---

* $f(x(u, v), y(u, v), z(u, v))$ is the function obtained by superposing $f(x, y, z)$ and $x=x(u, v)$, $y=y(u, v)$, $z=z(u, v)$. By the theorem on the continuity of a complex function this function is continuous in Ω.

in the bounded closed domain Ω and by the property of integrability of the function $f(x(u, v), y(u, v), z(u, v))\sqrt{EG-F^2}$ continuous in Ω.

We find $\delta > 0$ for $\delta^* > 0$ so that any subdivision of Φ with piecewise smooth curves into a finite number of parts Φ_i whose sizes are less than δ has a corresponding subdivision of Ω into a finite number of parts Ω_i whose sizes are less than δ^*. The possibility of a choice of such δ is guaranteed by the fact that Φ is a homeomorphic mapping of Ω and therefore corresponding to each subdivision of Φ with piecewise smooth curves into a finite number of parts Φ_i is a subdivision of Ω with piecewise smooth curves into a finite number of parts Ω_i. If the maximum size of the parts Φ_i tends to zero, then so does the maximum size of the parts Ω_i.

Now consider subdivision of Φ with piecewise smooth curves into a finite number of parts Φ_i whose maximum size Δ satisfies $\Delta < \delta$, where $\delta > 0$ is chosen for δ^* in the way shown above. Form for this subdivision a sum $I\{\Phi_i, M_i\}$ using (5.19). Since the area σ_i of a part Φ_i is $\iint\limits_{\Omega_i}\sqrt{EG-F^2}\,du\,dv$, denoting the coordinates of a point M_i in the part Φ_i by $(x(u_i, v_i), y(u_i, v_i), z(u_i, v_i))$ we get

$$I(\Phi_i, M_i) =$$
$$= \sum_i f(x(u_i, v_i), y(u_i, v_i), z(u_i, v_i)) \iint\limits_{\Omega_i}\sqrt{EG-F^2}\,du\,dv.$$

Using the mean value theorem for the integrals on the right of the last relation we clearly may transform that relation as follows:

$$I\{\Phi_i, M_i\} - \iint\limits_{\Omega} f(x(u, v), y(u, v), z(u, v))\sqrt{EG-F^2}\,du\,dv =$$
$$= \Big[\sum_i f(x(u_i, v_i), y(u_i, v_i), z(u_i, v_i)) \times$$
$$\times \sqrt{E(u_i, \bar{v}_i)G(u_i, v_i) - F^2(u_i, \bar{v}_i)}\,\sigma_i^* -$$
$$- \iint\limits_{\Omega} f(x(u, v), y(u, v), z(u, v))\sqrt{EG-F^2}\,du\,dv\Big] +$$
$$+ \sum_i f(x(u_i, v_i), y(u_i, v_i), z(u_i, v_i)) \times$$
$$\times \Big[\sqrt{E(\tilde{u}_i, \tilde{v}_i)G(\tilde{u}_i, \tilde{v}_i) - F^2(\tilde{u}_i, \tilde{v}_i)} -$$
$$- \sqrt{E(u_i, v_i)G(u_i, v_i) - F^2(u_i, v_i)}\Big]\sigma_i^*.$$

From the last equation, via inequalities (5.28) and (5.29), we easily obtain inequality (5.27). The proof of the theorem is complete.

Remark 1. To compute the surface integral of the second kind $\iint\limits_{\Phi} f(x, y, z) \cos Z \, d\sigma$ once a definite side of Φ is chosen, we can clearly use the following formula:

$$\iint\limits_{\Phi} f(x, y, z) \cos Z \, d\sigma =$$
$$= \iint\limits_{\Omega} f(x(u, v), y(u, v), z(u, v)) \sqrt{EG - F^2} \cos Z \, du \, dv. \qquad (5.30)$$

Similar formulas are true for the other two surface integrals of the second kind.

Remark 2. Let a surface Φ be the graph of the function $z = z(x, y)$ belonging to its domain of definition, D, to the class C^1. Choose on Φ the side for which the unit normal vector $\boldsymbol{n}(M)$ of Φ makes with the Oz axis an acute angle. In that case $\cos Z = \frac{1}{\sqrt{1+p^2+q^2}}$, where $p = \frac{\partial z}{\partial x}$, $q = \frac{\partial z}{\partial y}$. Suppose on Φ a continuous function $R(x, y, z)$ is given. Considering then that x and y are taken to be the parameters u and v on Φ (Φ is defined by the parametric equations $x = x$, $y = y$, $z = z(x, y)$) and $\sqrt{EG - F^2} = \sqrt{1 + p^2 + q^2}$, we may rewrite formula (5.30) as

$$\iint\limits_{\Phi} R(x, y, z) \cos Z \, d\sigma = \iint\limits_{D} R(x, y, z(x, y)) \sqrt{1+p^2+q^2} \times$$
$$\times \frac{1}{\sqrt{1+p^2+q^2}} \, dx \, dy = \iint\limits_{D} R(x, y, z(x, y)) \, dx \, dy.$$

This remark explains why we have the following notation for the surface integral of the second kind:

$$\iint\limits_{\Phi} R(x, y, z) \cos Z \, d\sigma = \iint\limits_{\Phi} R(x, y, z) \, dx \, dy. \qquad (5.31)$$

Note that (5.31) is used also in the case where Φ is not the graph of the function $z = z(x, y)$.

We shall consider surface integrals of the second kind having the following form:

$$\iint\limits_{\Phi} (P \cos X + Q \cos Y + R \cos Z) \, d\sigma.$$

Such integrals will also be designated as follows:

$$\iint_{\Phi} P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy.$$

Remark 3. The concepts of surface integrals of the first and the second kind can naturally be extended to the case where Φ is a piecewise smooth surface. For such surfaces clearly the existence theorem proved in this subsection is also true.

5.3.3. Surface integrals of the second kind independent of the choice of Cartesian coordinate system. From the definition of the surface integrals of the first and the second kind it follows that the integral of the first kind is independent of the choice of Cartesian coordinate system in space, whereas the integrals of the second kind are dependent of its choice, for changing the coordinate system changes the values of the cosines of the angles which the normal $n(M)$ makes with the coordinate axes.

In the case where a vector function is given on the surface we can show a more general approach to the notion of surface integral of the second kind which allows us to speak in a sense of the independence of the value of the integral of the choice of Cartesian coordinate system in space.

So let Φ be a smooth bounded complete two-sided surface on which a continuous vector function $\mathbf{r}(M)$ is given. Choose a definite side on Φ and denote by $\mathbf{n}(M)$ the vector field of unit normals to Φ.

Obviously the scalar product $\mathbf{r}(M)\,\mathbf{n}(M)$ is a continuous scalar function given on Φ and is therefore independent of the choice of Cartesian coordinate system in space. Consequently, the surface integral of the first kind of that function

$$\iint_{\Phi} \mathbf{r}(M)\,\mathbf{n}(M)\,d\sigma$$

is independent of the choice of Cartesian coordinate system in space. Use the coordinate notation of the scalar product $\mathbf{r}(M)\,\mathbf{n}(M)$, assuming that the vector $\mathbf{r}(M)$ has the coordinates P, Q, R. Since the coordinates of $\mathbf{n}(M)$ are $\cos X$, $\cos Y$, $\cos Z$, we have

$$\mathbf{r}(M)\,\mathbf{n}(M) = P\cos X + Q\cos Y + R\cos Z$$

and therefore

$$\iint_{\Phi} \mathbf{r}(M)\,\mathbf{n}(M)\,d\sigma = \iint_{\Phi} (P\cos X + Q\cos Y + R\cos Z)\,d\sigma.$$

The integral on the right of the last equation is the sum of the three surface integrals of the second kind and is usually called *the general*

surface integral of the second kind. Consequently, $\iint\limits_{\Phi} \mathbf{r}(M)\,\mathbf{n}(M)\,d\sigma$ may also be called the general surface integral of the second kind.

Remark 1. If on the surface Φ the three scalar functions P, Q, and R are given, the integral $\iint\limits_{\Phi} (P\cos X + Q\cos Y + R\cos Z)\,d\sigma$ may be written in the form invariant (independent) of coordinate system, assuming P, Q, and R to be the coordinates of some vector function $\mathbf{r}(M)$ given on the surface and writing the integral in the form $\iint\limits_{\Phi} \mathbf{r}(M)\,\mathbf{n}(M)\,d\sigma$. Note that thereby we impose a certain law of transforming the integrand in transition to a new Cartesian coordinate system. In this case we obtain new coordinates of the vector $\mathbf{r}(M)$ which are calculated by the well-known rule of analytic geometry. However, this invariant form of notation for the surface integral is very conveniently used in various applications.

Remark 2. Note that the general surface integral of the second kind $\iint\limits_{\Phi} \mathbf{r}(M)\,\mathbf{n}(M)\,d\sigma$ is numerically equal to the quantity called by physicists the *flux* of the vector $\mathbf{r}(M)$ through the surface Φ.

CHAPTER 6

BASIC FIELD THEORY OPERATIONS

This chapter considers scalar and vector fields. It investigates basic operations of field theory.

6.1. TRANSFORMATIONS OF BASES AND COORDINATES. INVARIANTS

6.1.1. Conjugate vector bases. Covariant and contravariant coordinates of vectors. Let r_i, $i = 1, 2, 3$, be a vector basis of a three-dimensional space* (for the plane, i assumes values of 1 and 2). The basis r^k, $k = 1, 2, 3$, is said to be *conjugate* to the basis r_i if**

$$r_i r^k = \delta_i^k = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases} \quad i, k = 1, 2, 3. \tag{6.1}$$

The symbol δ_i^k is called the Kronecker symbol***.

The question arises as to the existence and uniqueness of the conjugate basis. The answer to this question is yes: *for a given basis r_i there exists a unique conjugate basis r^k.*

We shall see for example that a vector r^1 is uniquely determined. According to (6.1) it is orthogonal to the vectors r_2 and r_3. This uniquely determines the line of action of r^1. Next from the condition $r_1 r^1 = 1$ the vector r^1 itself is uniquely determined. Similarly, the vectors r^2 and r^3 are uniquely constructed. To show that r^1, r^2, r^3 form a basis, it suffices to prove that $r^1 r^2 r^3 \neq 0$. By the theorem on the product of determinants

$$(r_1 r_2 r_3)(r^1 r^2 r^3) = \begin{vmatrix} r_1 r^1 & r_1 r^2 & r_1 r^3 \\ r_2 r^1 & r_2 r^2 & r_2 r^3 \\ r_3 r^1 & r_3 r^2 & r_3 r^3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1. \tag{6.2}$$

* Recall that r_1, r_2, r_3 form a basis if they are noncoplanar, i.e. if their triple product $r_1 r_2 r_3$ is nonzero.

** Throughout this chapter ab denotes the scalar product of vectors a and b, abc denotes the triple product of vectors a, b, and c, and $[ab]$ symbolizes the vector product of a and b.

*** Leopold Kronecker (1823-1891) is a German mathematician.

Since $r_1r_2r_3 \neq 0$ (r_1, r_2, r_3 form a basis), it follows from (6.2) that $r^1r^2r^3 \neq 0$ either.

Remark 1. If a basis r_i is orthonormal, then the conjugate basis r^k coincides with the given basis r_i.

It is easy to see that the vectors r^k of the conjugate basis in three-dimensional space can be found using the relations

$$r^1 = \frac{[r_2r_3]}{r_1r_2r_3}, \quad r^2 = \frac{[r_3r_1]}{r_1r_2r_3}, \quad r^3 = \frac{[r_1r_2]}{r_1r_2r_3}.$$

Let r_i, r^k be conjugate bases, and let x be an arbitrary vector. Expanding x with respect to the basis vectors we get

$$x = x_1r^1 + x_2r^2 + x_3r^3, \quad x = x^1r_1 + x^2r_2 + x^3r_3. \tag{6.3}$$

The numbers x_1, x_2, x_3 are called *covariant* coordinates of the vector x, and x^1, x^2, x^3 are the *contravariant* coordinates of x. These terms are made clear in Section 6.1.2.

To abbreviate the notation of formulas containing terms of the same type (relations (6.3) may serve as an example of such formulas), in what follows we shall employ the following convention. Suppose there is an expression made up of multipliers. If the expression has two identical literal indices, one superscript and the other subscript, it is considered that summation is taken with respect to these indices: the indices are successively assigned the values of 1, 2, 3 and the resulting terms are then added up. For example,

$$x_ir^i = x_1r^1 + x_2r^2 + x_3r^3, \quad \delta_i^i = \delta_1^1 + \delta_2^2 + \delta_3^3,$$

$$\begin{aligned} g_{ik}x^ix^k &= (g_{1k}x^1x^k) + (g_{2k}x^2x^k) + (g_{3k}x^3x^k) = \\ &= (g_{11}x^1x^1 + g_{12}x^1x^2 + g_{13}x^1x^3) + \\ &+ (g_{21}x^2x^1 + g_{22}x^2x^2 + g_{23}x^2x^3) + (g_{31}x^3x^1 + g_{32}x^3x^2 + g_{33}x^3x^3). \end{aligned}$$

Using the summation convention formulas (6.3) can be written in the following compact manner:

$$x = x_ir^i, \quad x = x^ir_i. \tag{6.4}$$

Remark 2. The superscripts and subscripts mentioned in the summation convention are usually referred to as dummy indices. It is clear that dummy indices may be denoted by any letters, the expression they occur in remaining unaffected. For instance, x_ir^i and x_kr^k represent the same expression.

Remark 3. All arguments in this subsection referred to the case of three-space. In the two-dimensional case the literal indices take the values of 1 and 2.

We obtain an expression for the covariant and contravariant coordinates of the vectors. To do this write the scalar product of the first of the equations (6.4) by r_k and that of the second by r^k.

Taking then into account relations (6.1) we find that

$$xr_k = x_i (r^i r_k) = x_i \delta^i_k = x_k,$$
$$xr^k = x^i (r_i r^k) = x^i \delta^k_i = x^k.$$

So

$$x_i = xr_i, \quad x^i = xr^i. \tag{6.5}$$

Using relations (6.5) we write formulas (6.4) as

$$x = (xr_i) r^i, \quad x = (xr^i) r_i. \tag{6.6}$$

Relations (6.6) are called the Gibbs* formulas. We again turn to the question of constructing conjugate bases.

Using formulas (6.6) we get

$$r^k = (r^k r^i) r_i, \quad r_k = (r_k r_i) r^i. \tag{6.7}$$

Denoting

$$g_{ki} = r_k r_i, \quad g^{ki} = r^k r^i, \tag{6.8}$$

we rewrite relations (6.7) as

$$r^k = g^{ki} r_i, \quad r_k = g_{ki} r^i. \tag{6.9}$$

So it suffices to know the matrix (g^{ki}) to construct the basis r^k from r_i and it suffices to know the matrix (g_{ki}) to construct r_k from r^i. We prove that these matrices are reciprocal. To do this we write the scalar product of the first of the equations (6.9) by r_j. Taking into account relations (6.1) we get

$$g^{ki} g_{ij} = \delta^k_j = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

These relations show that (g^{ki}) and (g_{ki}) are reciprocal. Since the elements of the reciprocal matrix can be calculated in terms of those of a given matrix, it is clear that using relations (6.9) solves the question of constructing conjugate bases.

6.1.2. Transformations of a basis and transformations of coordinates. Let r_i and r^i, $i = 1, 2, 3$, be conjugate bases, and let $r_{i'}$ and $r^{i'}$ be new conjugate bases.

Using the summation convention we write formulas for transformation of basis vectors. We have:

(1) formulas for transition from the basis r_i to a new basis $r_{i'}$ and reverse formulas:

$$r_{i'} = b^i_{i'} r_i, \quad r_i = b^{i'}_i r_{i'}, \quad i, i' = 1, 2, 3; \tag{6.10}$$

---

* Josiah Willard Gibbs (1839-1903) is an American theoretical physicist.

(2) formulas for transition from the basis r^i to a new basis $r^{i'}$ and reverse formulas

$$r^{i'}=\tilde{b}_i^{i'}r^i, \quad r^i=\tilde{b}_{i'}^i r^{i'}, \quad i,\ i'=1,\ 2,\ 3. \tag{6.11}$$

Since the transformations (6.10) are reciprocal, so are the matrices $(b_{i'}^i)$, and $(b_i^{i'})$. For similar reasons $(\tilde{b}_i^{i'})$ and $(\tilde{b}_{i'}^i)$ are also reciprocal.

We prove that $(b_{i'}^i)$ and $(\tilde{b}_{i'}^i)$ coincide. This will prove that so do $(b_i^{i'})$ and $(\tilde{b}_i^{i'})$. To begin with, we obtain the scalar product of the first of the equations (6.10) by r^k and that of the second of the equations (6.11) by $r_{k'}$. Taking then into account relations (6.1) we find

$$r_{i'}r^k=b_{i'}^i(r_ir^k)=b_{i'}^i\delta_i^k=b_{i'}^k,$$

$$r^ir_{k'}=\tilde{b}_{i'}^i(r^{i'}r_{k'})=\tilde{b}_{i'}^i\delta_{k'}^{i'}=\tilde{b}_{k'}^i.$$

From these relations we get

$$b_{i'}^i=r_{i'}r^i, \tag{6.12}$$

$$\tilde{b}_{i'}^i=r_{i'}r^i. \tag{6.13}$$

Since the right-hand sides of relations (6.12) and (6.13) are equal, so are the left-hand sides. In other words $b_{i'}^i=\tilde{b}_{i'}^i$ and this just means that the matrices $(b_{i'}^i)$ and $(\tilde{b}_{i'}^i)$ coincide. Note that the elements $b_{i'}$ of $(b_{i'}^i)$ can be calculated from formulas (6.12).

We may now say that to change from a basis r_i, r^i to $r_{i'}$, $r^{i'}$ it is sufficient to know only the matrix $(b_{i'}^i)$ of transition from the basis r_i to $r_{i'}$ (the matrix $(b_i^{i'})$ is calculated from $(b_{i'}^i)$).

Here is a complete list of formulas for transformations of basis vectors:

$$\left.\begin{aligned} r_{i'}&=b_{i'}^ir_i, & r_i&=b_i^{i'}r_{i'},\\ r^{i'}&=b_i^{i'}r^i, & r^i&=b_{i'}^ir^{i'}. \end{aligned}\right\} \tag{6.14}$$

We proceed to derive formulas for transforming the coordinates of a vector when changing to a new basis.

Let $x_{i'}$ be the covariant coordinates of x in the basis $r_{i'}$, $r^{i'}$. According to (6.5) we then have

$$x_{i'}=xr_{i'}.$$

Substituting into the right-hand side of this relation the expression for $r_{i'}$ from formulas (6.14) we find

$$x_{i'}=x(b_{i'}^ir_i)=b_{i'}^i(xr_i)=b_{i'}^ix_i.$$

So formulas for transforming the covariant coordinates of a vector in changing to a new basis are the form

$$x_{i'} = b^{i}_{i'} x_{i}. \qquad (6.14')$$

We see that in changing to a new basis the covariant coordinates of the vector x are transformed using the matrix $(b^{i}_{i'})$ of direct change from the old basis to a new one. This concordance of transformations explains the term "covariant* coordinates of a vector". Substituting into the right-hand side of $x^{i'} = xr^{i'}$ the expression for $r^{i'}$ in (6.14) we obtain the following formulas after transformations:

$$x^{i'} = b^{i'}_{i} x^{i}. \qquad (6.15)$$

We see that in changing to a new basis the contravariant coordinates of x are transformed using the matrix $(b^{i'}_{i})$ of inversion from the new basis to the old one. This discordance of transformations explains why the term "contravariant** coordinates of a vector" is used.

6.1.3. Invariants of a linear operator. The divergence and curl of a linear operator. *Invariants* are expressions independent of the choice of basis. For instance, the value of the scalar function at a given point is an invariant. An invariant is an object vector independent of the choice of basis. A scalar product of vectors is also an invariant.

Here we shall discuss some invariants of a linear operator. Let A be an arbitrary linear operator defined on vectors of a three-dimensional Euclidean space (i.e. $A(\alpha x + \beta y) = \alpha Ax + \beta Ay$ for any vectors x and y and any real numbers α and β). We prove that the expression

$$r^{i} A r_{i} = r_{i} A r^{i} \; *** \qquad (6.16)$$

is an invariant.

We want to prove that in changing to another basis $r_{i'}$, $r^{i'}$

$$r^{i} A r_{i} = r^{i'} A r_{i'}. \qquad (6.17)$$

Let $r_{i'}$, $r^{i'}$ be a new basis and let $(b^{i}_{i'})$ be the matrix of transition from the basis r_{i}, r^{i} to $r_{i'}$, $r^{i'}$. We have

$$r_{i} = b^{i'}_{i} r_{i'}, \quad r^{i} = b^{i}_{k'} r^{k'}.$$

---

* Covariant means concordantly changing.

** Contravariant means changing in an opposite way.

*** The validity of the equation $r^{i} A r_{i} = r_{i} A r^{i}$ can be seen from the following. According to (6.9) $r^{l} = g^{lk} r_{k}$, $r_{i} = g_{il} r^{l}$. Considering that the matrices (g^{ik}) and (g_{il}) are reciprocal and symmetrical we therefore get

$$r^{i} A r_{i} = g^{ik} g_{il} r_{k} A r^{l} = \delta^{k}_{l} r_{k} A r^{l} = r_{k} A r^{k} = r_{i} A r^{i}.$$

Substituting these values for r_i and r^i in the expression r^iAr_i we get

$$r^iAr_i = (b_i^{i'}b_{k'}^i)\, r^{k'}Ar_{i'}. \qquad (6.18)$$

Since $(b_i^{i'}b_{k'}^i) = \delta_{k'}^{i'}$, from (6.18) we get

$$r^iAr_i = \delta_{k'}^{i'}r^{k'}Ar_{i'} = r^{i'}Ar_{i'}.$$

Equation (6.17) is thus proved, and hence so is the invariance of the expression r^iAr_i.

The invariant r^iAr_i of a linear operator A will be called the *divergence* of the operator and designated div A. Thus

$$\operatorname{div} A = r^iAr_i = r_iAr^i. \qquad (6.19)$$

Remark. In a given basis r_i, r^i a linear operator may be given using a matrix called the matrix of the linear operator. It is the matrix of the coefficients a_i^k of expansion of vectors Ar_i with respect to the basis r_k (it is possible of course to consider the matrix of the coefficients of expansion of Ar^i with respect to r^k):

$$Ar_i = a_i^kr_k; \quad a_i^j = r^jAr_i. \qquad (6.20)$$

The divergence of the matrix of A may be expressed in terms of the elements of the matrix (a_i^k). Namely

$$\operatorname{div} A = a_i^i = a_1^1 + a_2^2 + a_3^3. \qquad (6.21)$$

To see that formula (6.21) is valid it is sufficient to substitute the expression (6.20) for Ar_i in the expression (6.19) for divergence and use the relation $r^ir_k = \delta_k^i$.

We prove that the expression

$$[r_iAr^i] = [r^iAr_i]\,^* \qquad (6.22)$$

is also an invariant. We must prove that in changing from the basis r_i, r^i to another basis $r_{i'}$, $r^{i'}$

$$[r_iAr^i] = [r_{i'}Ar^{i'}]. \qquad (6.23)$$

Let $r_{i'}$, $r^{i'}$ be a new basis and let $(b_{i'}^i)$ be the matrix of transition from the basis r_i, r^i to $r_{i'}$, $r^{i'}$. We have

$$r_i = b_i^{i'}r_{i'}, \quad r^i = b_{k'}^ir^{k'}.$$

---

* The validity of $[r_iAr^i] = [r^iAr_i]$ can be seen from the following. According to (6.9) $r^i = g^{ik}r_k$, $r_i = g_{il}r^l$. Using the reciprocity and symmetry of (g^{ik}) and (g_{il}) we therefore get

$$[r^iAr_i] = g^{ik}g_{il}\,[r_kAr^l] = \delta_l^k\,[r_kAr^l] = [r_kAr^k] = [r_iAr^i].$$

Substituting these values for r_i and r^i in $[r_iAr^i]$ we get

$$[r_iAr^i] = (b_i^{i'} b_{k'}^i)\, [r_{i'}Ar^{k'}]. \qquad (6.24)$$

Since $(b_i^{i'}\ b_{j'}^i) = \delta_{k'}^{i'}$, from (6.24) we get

$$[r_iAr^i] = \delta_{k'}^{i'}\, [r_{i'}Ar^{k'}] = [r_{i'}Ar^{i'}].$$

Equation (6.23) is thus proved, and hence so is the invariance of $[r_iAr^i]$.

The invariant $[r_iAr^i]$ of a linear operator A will be called the *curl* of the operator and designated curl A. Thus

$$\operatorname{curl} A = [r_iAr^i] = [r_1Ar^1] + [r_2Ar^2] + [r_3Ar^3]. \qquad (6.25)$$

We show the expression for the divergence and curl of the linear operator A for the case of the *orthonormal* basis i, j, k. Since in this case the conjugate basis coincides with the given basis, by formulas (6.20) the elements a_{ij} of the matrix of A can be found from the formulas

$$\left.\begin{aligned} a_{11} &= iAi, & a_{12} &= iAj, & a_{13} &= iAk,\\ a_{21} &= jAi, & a_{22} &= jAj, & a_{23} &= jAk,\\ a_{31} &= kAi, & a_{32} &= kAj, & a_{33} &= kAk, \end{aligned}\right\} \qquad (6.26)$$

(in contrast to the general case we have denoted the elements of the matrix of A by a_{ml} instead of a_l^m).

For the divergence of A we obtain the following expression:

$$\operatorname{div} A = \sum_{i=1}^{3} a_{ii} = a_{11} + a_{22} + a_{33} = iAi + jAj + kAk. \qquad (6.27)$$

We find the expression for the curl of A. Since in the case of the orthonormal basis conjugate bases coincide, from (6.25) we get

$$\operatorname{curl} A = [iAi] + [jAj] + [kAk]. \qquad (6.28)$$

We evaluate the first vector product $[iAi]$. Since $Ai = a_{11}i + a_{21}j + a_{31}k$, we have

$$[iAi] = a_{11}\,[ii] + a_{21}\,[ij] + a_{31}\,[ik] = -a_{31}j + a_{21}k.$$

Quite similarly we obtain the formulas

$$[jAj] = a_{32}i - a_{12}k, \quad [kAk] = -a_{23}i + a_{13}j.$$

Using these and relations (6.28) for curl A we find

$$\operatorname{curl} A = (a_{32} - a_{23})\, i + (a_{13} - a_{31})\, j + (a_{21} - a_{12})\, k. \qquad (6.29)$$

6.2. THE SCALAR FIELD AND THE VECTOR FIELD. BASIC CONCEPTS AND OPERATIONS

6.2.1. The scalar field and the vector field. Let Ω be a domain in the plane or in space.

A scalar field is said to be given in Ω if associated with each point M of Ω by a certain law is some number $u(M)$.

Note that the concepts of scalar field and function defined in Ω coincide. The following terminology is commonly employed: *a scalar field is given using a function $u(M)$.*

The notion of vector field is introduced quite similarly: *if associated with each point M in Ω by a certain law is some vector $\boldsymbol{p}(M)$, then a vector field is said to be given* in Ω. We shall employ the terminology: *a vector field is given using a vector function $\boldsymbol{p}(M)$.*

The temperature field inside a heated body, the field of mass density are examples of scalar fields. The velocity field of a steady-state liquid flow, magnetic intensity field are examples of vector fields.

6.2.2. Differentiable scalar fields. The gradient of a scalar field. The directional derivative. We have already noted that the notions of scalar field $u(M)$ in a domain Ω and function defined in that field coincide. We may therefore define the differentiability of a scalar field as the differentiability of the function giving that field. For convenience we shall formulate the notion of differentiability of the field employing terminology somewhat different from the usual terminology.

A *linear form* $f(\Delta\boldsymbol{r})$ in a vector $\Delta\boldsymbol{r}$ is the scalar product of that vector by some $\Delta\boldsymbol{r}$-independent vector $\boldsymbol{g}$. We shall also use the notation:

$\rho = \rho(M, M')$ is the distance between points M and M',

$\Delta\boldsymbol{r} = \overline{MM'}$ is the vector connecting M and M',

$\Delta u = u(M') - u(M)$ is the increment of the field at M.

We give the following *definition.*

Definition 1. A scalar field $u(M)$ is said to be **differentiable** *at a point M of a domain Ω if the increment of the field Δu at M may be represented in the following form:*

$$\Delta u = f(\Delta\boldsymbol{r}) + o(\rho), \tag{6.30}$$

where $f(\Delta\boldsymbol{r})$ is a linear form in a vector $\Delta\boldsymbol{r}$.

Relation (6.30) will be called the *differentiability condition* of the field $u(M)$ at a point M.

Remark 1. Since a linear form $f(\Delta\boldsymbol{r})$ is a scalar product $\boldsymbol{g}\cdot\Delta\boldsymbol{r}$, where $\boldsymbol{g}$ is a $\Delta\boldsymbol{r}$-independent vector, the differentiability condition of $u(M)$ at M may be written as

$$\Delta u = \boldsymbol{g}\cdot\Delta\boldsymbol{r} + o(\rho). \tag{6.31}$$

We prove that if a scalar field $u(M)$ is differentiable at a point M, then the representation (6.30) (or (6.31)) for the increment Δu of that field at M is unique. Let

$$\Delta u = g\cdot\Delta r + o_1(\rho) \text{ and } \Delta u = h\cdot\Delta r + o_2(\rho) \tag{6.32}$$

be two representations of the increment Δu at M. From formulas (6.32) for $\Delta r \neq 0$ we obtain the relation

$$(g-h)\,e = \frac{o(\rho)}{|\Delta r|}, \tag{6.33}$$

where $e = \frac{\Delta r}{|\Delta r|}$ is the unit vector and $o(\rho) = o_2(\rho) - o_1(\rho)$. Since $\frac{o(\rho)}{|\Delta r|} = \frac{o(\rho)}{\rho}$ is an infinitesimal as $\rho \to 0$, it follows from (6.33) that $(g-h)\,e = 0$ for any e, i.e. $g = h$. The uniqueness of the representation (6.30) is thus proved.

We shall say that *a scalar field $u(M)$ given in a domain Ω is differentiable in Ω if it is differentiable at each point of Ω.*

Definition 2. The gradient *at a point M of a scalar field $u(M)$ differentiable at that point is a vector g defined by relation* (6.31).

The gradient of a scalar field is designated grad u.

Remark 2. The above definition of the differentiability of a scalar field is convenient in that it has an invariant character independent of the choice of coordinate system. *The gradient of a scalar field is* therefore *an invariant of that field.*

Remark 3. Note the following important fact: *if a scalar field $u(M)$ given in a domain Ω is differentiable in that domain, then the gradient* grad u *of the field is defined at each point of Ω and is obviously a vector field given in Ω.*

Remark 4. For the scalar field we introduce the concept of *level surface* (*level curve* for the plane field) which is a set of points on which the values of the field $u(M)$ are the same. The gradient of the field at a point M is orthogonal to the level surface at that point. The reader will easily see for himself that this remark is valid.

Using the symbol grad u for the gradient of a scalar field we rewrite relation (6.31) in the following form:

$$\Delta u = \operatorname{grad} u\cdot\Delta r + o(\rho). \tag{6.34}$$

Note that the term grad $u\cdot\Delta r$ is usually called the *differential du* of the scalar field. Thus

$$du = \operatorname{grad} u\cdot\Delta r. \tag{6.35}$$

Let us agree to employ the term differential dr for the increment Δr of the radius vector $r = \overline{OM}$, $\Delta r = \overline{OM'} - \overline{OM}$. Formula (6.35) for the differential du of a scalar field may be written as

$$du = \operatorname{grad} u\cdot dr. \tag{6.36}$$

Let two differentiable fields u (M) and v (M) be given in Ω. The following relations are true:

$$\left.\begin{aligned} &\operatorname{grad}(u \pm v) = \operatorname{grad} u \pm \operatorname{grad} v, \\ &\operatorname{grad}(uv) = u \operatorname{grad} v + v \operatorname{grad} u, \\ &\operatorname{grad}\left(\frac{u}{v}\right) = \frac{v \operatorname{grad} u - u \operatorname{grad} v}{v^2} \quad (\text{for } v \neq 0). \end{aligned}\right\} \tag{6.37}$$

If F is a differentiable function, then

$$\operatorname{grad} F(u) = F'(u) \operatorname{grad} u. \tag{6.38}$$

The derivations of formulas (6.37) and (6.38) are of the same type. As an illustration, we show the validity of the second of the formulas (6.37). Using formula (6.34) and the continuity of the function u (M) we have

$$\begin{aligned} \Delta(uv) &= u(M')\,v(M') - u(M)\,v(M) = \\ &= u(M')\,\Delta v + v(M)\,\Delta u = \\ &= (u(M) \operatorname{grad} v + v(M) \operatorname{grad} u)\,\Delta r + o(\rho). \end{aligned}$$

From these relations it follows that the increment Δ (uv) may be represented by the form (6.31). Therefore uv is a differentiable function and grad (uv) = u grad v + v grad u. The second of the formulas (6.37) is thus proved.

We introduce the concept of *directional derivative* for the scalar field.

Let a field u (M) be given in Ω, let M be some point of Ω, and let e be the unit vector indicating the direction at M. Also suppose that M' is any point in Ω other than M and such that the vector $\overline{MM'}$ is collinear with e. The distance between M and M' is denoted by ρ.

If there is a limit

$$\lim_{\rho \to 0} \frac{\Delta u}{\rho}$$

($\Delta u = u(M') - u(M)$), *then this limit is said to be the derivative of the field u at M with respect to the direction e and designated* $\frac{\partial u}{\partial e}$. Thus

$$\frac{\partial u}{\partial e} = \lim_{\rho \to 0} \frac{\Delta u}{\rho}. \tag{6,39}$$

The following statement is true.

Let u (M) be differentiable at a point M. Then the derivative $\frac{\partial u}{\partial e}$ of u at M with respect to any direction e exists and can be found from the

formula

$$\frac{\partial u}{\partial e} = \operatorname{grad} u \cdot e. \tag{6.40}$$

We prove this statement. Let e be any fixed direction and let the point M' be taken so that the vector $\Delta r = \overline{MM'}$ is collinear with e. It is clear that $\Delta r = \rho e$. Substituting the value of Δr in relation (6.34) we find

$$\Delta u = (\operatorname{grad} u \cdot e)\, \rho + o(\rho).$$

From this we obtain the formula

$$\frac{\Delta u}{\rho} = \operatorname{grad} u \cdot e + \frac{o(\rho)}{\rho}. \tag{6.41}$$

Relations (6.39) and (6.41) yield formula (6.40). This proves the statement.

We find the expression for the gradient of a differentiable scalar field assuming that in space an orthonormal basis i, j, k is chosen referred to a Cartesian coordinate system $Oxyz$. Since $\operatorname{grad} u = i(\operatorname{grad} u \cdot i) + j(\operatorname{grad} u \cdot j) + k(\operatorname{grad} u \cdot k)$ and $\frac{\partial u}{\partial i} = \frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial j} = \frac{\partial u}{\partial y}$, $\frac{\partial u}{\partial k} = \frac{\partial u}{\partial z}$, using relations (6.40) we get

$$\operatorname{grad} u = i\,\frac{\partial u}{\partial x} + j\,\frac{\partial u}{\partial y} + k\,\frac{\partial u}{\partial z}.$$

Using the expressions (6.40) for the directional derivative gives the following graphical picture of distribution of the values of directional derivatives of the field $u(M)$ in a plane domain Ω at a given point M. Let $\operatorname{grad} u \neq 0$ (if $\operatorname{grad} u = 0$, then it follows from (6.40) that $\frac{\partial u}{\partial e} = 0$ for any e). On the vector grad u as diameter (Fig. 6.1) we construct a circle C. We also construct a circle C^* as large as C and touching it at M. Let e be an arbitrary direction. Draw through M a half-line in the direction of the vector e. If the half-line touches C and C^*, then $\frac{\partial u}{\partial e} = 0$ (the vector e is orthogonal to grad u). If, however, the half-line intersects C or C^* in a point N, then $\frac{\partial u}{\partial e}$ is equal to the length of MN taken with the plus sign when N is on C and with the minus sign when N is on C^*. For a field in space the circles C and C^* must be replaced with spheres.

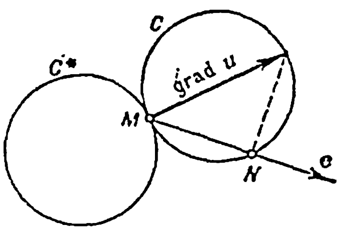


Fig. 6.1

6.2.3. Differentiable vector fields. The divergence and curl of a vector field. The directional derivative of a vector field. Let a vector field $p(M)$ be given in a domain Ω of a three-dimensional Euclidean space. In what follows we use the notation: $\Delta r = \overline{MM'}$, $\Delta p = p(M') - p(M)$.

We give the following definition.

Definition 3. A vector field $p(M)$ is said to be **differentiable** *at a point M of Ω if the increment of the field Δp at M may be represented in the following form:*

$$\Delta p = A\Delta r + o(|\Delta r|), \tag{6.42}$$

where A is a linear operator independent of Δr (independent of the choice of point M').

Relation (6.42) will be called *the differentiability conditions of a field $p(M)$ at a point M.*

We prove that if $p(M)$ is differentiable at M, then the representation (6.42) for the increment Δp of $p(M)$ at M is unique.

Let

$$\Delta p = A\Delta r + o_1(|\Delta r|) \text{ and } \Delta p = B\Delta r + o_2(|\Delta r|) \tag{6.43}$$

be two representations of Δp at M. For $\Delta r \neq 0$ formulas (6.43) give

$$(A-B)e = \frac{o(|\Delta r|)}{|\Delta r|}, \tag{6.44}$$

where $e = \frac{\Delta r}{|\Delta r|}$ is the unit vector, $o(|\Delta r|) = o_2(|\Delta r|) - o_1(|\Delta r|)$. Since $\frac{o(|\Delta r|)}{|\Delta r|}$ is an infinitesimal vector as $\Delta r \to 0$ and e is an arbitrary unit vector, it follows from (6.44) that $(A-B)e = 0$ for any e, i.e. $A = B$. This proves the uniqueness of the representation (6.42).

We shall say that *a vector field $p(M)$ given in a domain Ω is differentiable in that domain if it is differentiable at each point of Ω.*

We introduce *the concept of directional derivative for a vector field $p(M)$.*

Let the field $p(M)$ be given in a domain Ω, let M be some point of Ω, and let e be the unit vector indicating the direction at M. Also suppose that M' is any point in Ω other than M and such that the vector $\overline{MM'}$ is collinear with e. The distance between M and M' is denoted by ρ.

If there is a limit

$$\lim_{\rho\to 0} \frac{\Delta p}{\rho}$$

($\Delta p = p(M') - p(M)$), *then this limit is said to be the derivative of $p(M)$ at M with respect to the direction e and designated* $\frac{\partial p}{\partial e}$.

Thus

$$\frac{\partial p}{\partial e} = \lim_{\rho \to 0} \frac{\Delta p}{\rho}. \qquad (6.45)$$

The following statement is true. Let a field $p(M)$ *be differentiable at a point* M *of* Ω. *Then the derivative* $\frac{\partial p}{\partial e}$ *of the field* p *at that point with respect to any direction* e *exists and can be found from the formula*

$$\frac{\partial p}{\partial e} = Ae, \qquad (6.46)$$

where A *is a linear operator defined by relation* (6.42).

We prove this statement. Let e be any fixed direction and let the point M' be taken so that the vector $\Delta r = \rho e$ and $|\Delta r| = \rho$. Substituting this value of Δr in relation (6.42) and using the properties of a linear operator we find

$$\Delta p = \rho Ae + o(\rho).$$

From this we obtain the formula

$$\frac{\Delta p}{\rho} = Ae + \frac{o(\rho)}{\rho}. \qquad (6.47)$$

Relations (6.45) and (6.47) yield formula (6.46). This proves the statement.

Let $p(M)$ be a field differentiable at a point M of Ω. Then

$$\Delta p = A\Delta r + o(|\Delta r|).$$

We find the matrix of a linear operator A for the case of the orthonormal basis i, j, k. We shall assume that this basis is referred to a Cartesian coordinate system $Oxyz$.

We denote by P, Q and R the coordinates of $p(M)$ in i, j, k. Obviously by (6.46)

$$\frac{\partial p}{\partial i} = \frac{\partial p}{\partial x} = Ai, \quad \frac{\partial p}{\partial j} = \frac{\partial p}{\partial y} = Aj, \quad \frac{\partial p}{\partial k} = \frac{\partial p}{\partial z} = Ak.$$

From these formulas and from relations (6.26) for the matrix of the coefficients of a linear operator in i, j, k it follows that the matrix $\dot{A}$ of the operator A in question has the form

$$\dot{A} = \begin{pmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} & \frac{\partial P}{\partial z} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} & \frac{\partial Q}{\partial z} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} & \frac{\partial R}{\partial z} \end{pmatrix}. \qquad (6.48)$$

We introduce the concept of *divergence* and *curl* of a vector field $p(M)$ differentiable in Ω, i.e. of a field such that its increment Δp

at each point M of Ω may be represented as

$$\Delta \boldsymbol{p} = A \Delta \boldsymbol{r} + o(|\Delta \boldsymbol{r}|),$$

with A in general changing in going from point to point in Ω. In other words, A is dependent on a point M and is of course independent of $\Delta \boldsymbol{r}$.

We shall call the divergence and curl of a linear operator A the divergence and curl of a field $\boldsymbol{p}(M)$ at a point M of a domain Ω. Thus by definition

$$\operatorname{div} \boldsymbol{p} = \operatorname{div} A, \quad \operatorname{curl} \boldsymbol{p} = \operatorname{curl} A. \tag{6.49}$$

Remark. Under our hypotheses about the differentiability of the field $\boldsymbol{p}(M)$ in Ω $\operatorname{div} \boldsymbol{p}$ and $\operatorname{curl} \boldsymbol{p}$ are defined at each point of Ω. Since these objects are invariants (independent of the choice of basis), clearly $\operatorname{div} \boldsymbol{p}$ is a scalar field and $\operatorname{curl} \boldsymbol{p}$ is a vector field in Ω.

We find the expressions for the divergence, curl and directional derivative of a differentiable vector field $\boldsymbol{p}(M)$ assuming that in space an orthonormal basis $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ is chosen referred to a Cartesian coordinate system $Oxyz$. We shall assume as above that $\boldsymbol{p}(M)$ has coordinates P, Q, R in $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$.

Since the matrix $\mathring{A}$ of a linear operator A is defined by relation (6.48) in the case under consideration and by definition $\operatorname{div} \boldsymbol{p} = \operatorname{div} A$, $\operatorname{curl} \boldsymbol{p} = \operatorname{curl} A$ (see (6.49)), from formulas (6.27) and (6.29) we get

$$\operatorname{div} \boldsymbol{p} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \tag{6.50}$$

$$\operatorname{curl} \boldsymbol{p} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) \boldsymbol{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) \boldsymbol{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) \boldsymbol{k}. \tag{6.51}$$

To compute the derivative of $\boldsymbol{p}(M)$ with respect to $\boldsymbol{e}$ we use formula (6.46) and the properties of the linear operator.

Let $\boldsymbol{e} = \boldsymbol{i} \cos\alpha + \boldsymbol{j} \cos\beta + \boldsymbol{k} \cos\gamma$*. Then by (6.46) we get

$$\begin{aligned}\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{e}} &= A\boldsymbol{e} = \cos\alpha\, A\boldsymbol{i} + \cos\beta\, A\boldsymbol{j} + \cos\gamma\, A\boldsymbol{k} = \\ &= \cos\alpha \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{i}} + \cos\beta \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{j}} + \cos\gamma \frac{\partial \boldsymbol{p}}{\partial \boldsymbol{k}} = \\ &= \cos\alpha \frac{\partial \boldsymbol{p}}{\partial x} + \cos\beta \frac{\partial \boldsymbol{p}}{\partial y} + \cos\gamma \frac{\partial \boldsymbol{p}}{\partial z}.\end{aligned}$$

Thus the derivative $\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{e}}$ can be computed either from the formula

$$\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{e}} = \cos\alpha \frac{\partial \boldsymbol{p}}{\partial x} + \cos\beta \frac{\partial \boldsymbol{p}}{\partial y} + \cos\gamma \frac{\partial \boldsymbol{p}}{\partial z}, \tag{6.52}$$

* Since $\boldsymbol{e}$ is a unit vector, its coordinates are of the form $\{\cos\alpha, \cos\beta, \cos\gamma\}$, where α, β and γ are the angles $\boldsymbol{e}$ makes with the Ox, Oy, and Oz axes respectively.

or, considering that P, Q, R are the coordinates of $p(M)$, from the formula

$$\frac{\partial p}{\partial e}=\left(\frac{\partial P}{\partial x}\cos\alpha+\frac{\partial P}{\partial y}\cos\beta+\frac{\partial P}{\partial z}\cos\gamma\right)i+$$
$$+\left(\frac{\partial Q}{\partial x}\cos\alpha+\frac{\partial Q}{\partial y}\cos\beta+\frac{\partial Q}{\partial z}\cos\gamma\right)j+$$
$$+\left(\frac{\partial R}{\partial x}\cos\alpha+\frac{\partial R}{\partial y}\cos\beta+\frac{\partial R}{\partial z}\cos\gamma\right)k. \tag{6.53}$$

6.2.4. Compositions of field theory operations. We shall assume that a scalar field $u(M)$ of the class C^2* and a vector field $p(M)$ of the class C^2 are given in a domain Ω of a Euclidean space E^3.

Under these hypotheses grad u is a differentiable vector field in Ω, div p is a differentiable scalar field and curl p is a differentiable vector field. Therefore the following compositions of operations are permissible:

curl grad u, div grad u, grad div p, div curl p, curl curl p.

We prove that

$$\text{curl grad } u=0,\quad \text{div curl } p=0. \tag{6.54}$$

To this end we compute curl grad u and div curl p in a Cartesian system. Since in this case the coordinates of grad u are $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial u}{\partial z}$, on the basis of formula (6.51) we have

$$\text{curl grad } u=\left(\frac{\partial^2 u}{\partial y\,\partial z}-\frac{\partial^2 u}{\partial z\,\partial y}\right)i+\left(\frac{\partial^2 u}{\partial z\,\partial x}-\frac{\partial^2 u}{\partial x\,\partial z}\right)j+$$
$$+\left(\frac{\partial^2 u}{\partial x\,\partial y}-\frac{\partial^2 u}{\partial y\,\partial x}\right)k=0.$$

Thus the first of the equations (6.54) is true for a Cartesian system. By the invariance of the expression for curl grad u, the first of the equations (6.54) is proved. We proceed to prove the second of the equations (6.54). We again turn to a Cartesian system. In this system, according to (6.51), a vector field curl p has coordinates $\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)$, $\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)$, $\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)$, where P, Q, R are the coordinates of a vector p. According to (6.50) the divergence of a vector field curl p in Cartesian coordinates is equal to the sum of the derivatives of the components of curl

* A function belongs to a class C^k in Ω if all of its partial derivatives of order k are continuous.

p with respect to like coordinates. Thus

$$\operatorname{div}\operatorname{curl} p = \frac{\partial}{\partial x}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) = 0.$$

The second of the equations (6.54) is thus true for a Cartesian system. By virtue of the invariance of the expression for div curl p, the second of the equations (6.54) is true in any coordinate system.

One of the basic composed operations of field theory is the operation div grad u. It is briefly designated Δu, the symbol Δ being usually called the Laplacian* (operator). Thus

$$\Delta u = \operatorname{div}\operatorname{grad} u. \tag{6.55}$$

We evaluale the Laplacian in a Cartesian system. In such a system a vector field grad u has coordinates $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}$. Using the expression (6.50) for the divergence of a vector field we get

$$\Delta u \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}. \tag{6.56}$$

The composed operations grad div p and curl curl p are connected by the relation

$$\operatorname{curl}\operatorname{curl} p = \operatorname{grad}\operatorname{div} p - \Delta p, \tag{6.57}$$

where Δp is a vector whose coordinates in the basis i, j, k are equal to ΔP, ΔQ, ΔR (P, Q, R are the coordinates of a vector field p in i, j, k). The reader can easily show on his own the validity of relation (6.57).

6.3. EXPRESSING BASIC FIELD THEORY OPERATIONS IN CURVILINEAR COORDINATES

6.3.1. Curvilinear coordinates. Let Ω be a domain of a Euclidean space E^3; let x, y, z be Cartesian coordinates in that space. Suppose further that $\tilde{\Omega}$ is a domain of a Euclidean space $\tilde{E}^3$ and that x^1, x^2, x^3 are Cartesian coordinates in $\tilde{E}^3$.

Consider a one-to-one and bicontinuous mapping of $\tilde{\Omega}$ onto Ω effected via the functions

$$x = x(x^1, x^2, x^3),\ y = y(x^1, x^2, x^3),\ z = z(x^1, x^2, x^3). \tag{6.58}$$

Using the above mapping we introduce in Ω curvilinear coordinates x^1, x^2, x^3. The meaning of the term is easy to see from the following

* Pierre Simon Laplace (1749-1827) is an outstanding French astronomer, mathematician, and physicist.

arguments. First, associated with each point $M(x, y, z)$ of Ω are three numbers x^1, x^2, x^3. More precisely, M is determined by a triple of numbers x^1, x^2, x^3. This explains the term "coordinates" of point M for numbers x^1, x^2, x^3. Secondly, if on the right-hand sides of relations (6.58) any two coordinates, for example x^2 and x^3, are fixed, then for a variable x^1 these relations define in Ω some curve different in general from a straight line. It is natural to call that curve *a coordinate curve* x^1, thus emphasizing that at the points of the curve only the coordinate x^1 changes. The coordinate curves x^2 and x^3 are defined quite similarly. In general the coordinate curves x^1, x^2, and x^3 are not straight lines. This explains the term "curvilinear coordinates".

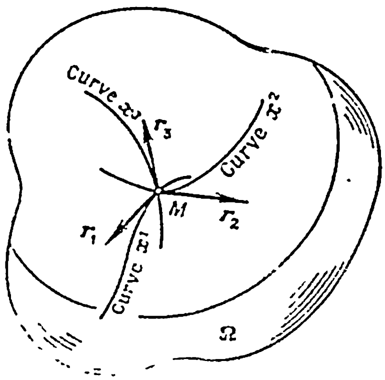


Fig. 6.2

We have shown that there are three coordinate curves x^1, x^2, x^3 passing through each M of Ω (Fig. 6.2). We construct at a point M a basis r_i, r^i associated in a natural way with the coordinate curves passing through that point. In doing so we use relations (6.58). Obviously the derivatives $\frac{\partial x}{\partial x^1}$, $\frac{\partial y}{\partial x^1}$, $\frac{\partial z}{\partial x^1}$, calculated at a point M are the coordinates of the tangent vector to x^1 at that point. We denote this vector by r_1. In a similar way we construct the tangent vectors r_2 and r_3 to the curves x^2 and x^3 respectively. Thus,

$$r_k = \left\{\frac{\partial x}{\partial x^k}, \frac{\partial y}{\partial x^k}, \frac{\partial z}{\partial x^k}\right\}, \quad k = 1, 2, 3. \tag{6.59}$$

For the vectors r_1, r_2, r_3 to form a basis we must require that they should be noncoplanar. A sufficient condition for this requirement to be met is obviously the condition that the Jacobian

$$\frac{\mathcal{D}(x, y, z)}{\mathcal{D}(x^1, x^2, x^3)} = \begin{vmatrix} \frac{\partial x}{\partial x^1} & \frac{\partial y}{\partial x^1} & \frac{\partial z}{\partial x^1} \\ \frac{\partial x}{\partial x^2} & \frac{\partial y}{\partial x^2} & \frac{\partial z}{\partial x^2} \\ \frac{\partial x}{\partial x^3} & \frac{\partial y}{\partial x^3} & \frac{\partial z}{\partial x^3} \end{vmatrix}$$

should be nonvanishing, for it is equal to the triple product of the vectors r_1, r_2, r_3. Using the constructed basis r_1, r_2, r_3 the conjugate basis r^1, r^2, r^3 is constructed in a standard way.

So, if curvilinear coordinates x^1, x^2, x^3 are introduced in a domain Ω, basis vectors $\boldsymbol{r}_i$, $\boldsymbol{r}^i$ are associated with each point M of Ω in a natural way. Consider the examples.

1°. *Cylindrical coordinate system.* It is introduced with the aid of the relations

$$x = \rho \cos \varphi, \quad y = \rho \sin \varphi, \quad z = z. \tag{6.60}$$

Thus $x^1 = \rho$, $x^2 = \varphi$, $x^3 = z$. It is known that the coordinates ρ, φ, z (or, equivalently, x^1, x^2, x^3) change within the following limits:

$$0 \leqslant \rho < +\infty, \quad 0 \leqslant \varphi < 2\pi, \quad -\infty < z < \infty.$$

These inequalities define in a Euclidean space $\tilde{E}^3$ with coordinates ρ, φ, z, (or x^1, x^2, x^3) an infinite domain $\tilde{\Omega}$ represented in Fig. 6.3.

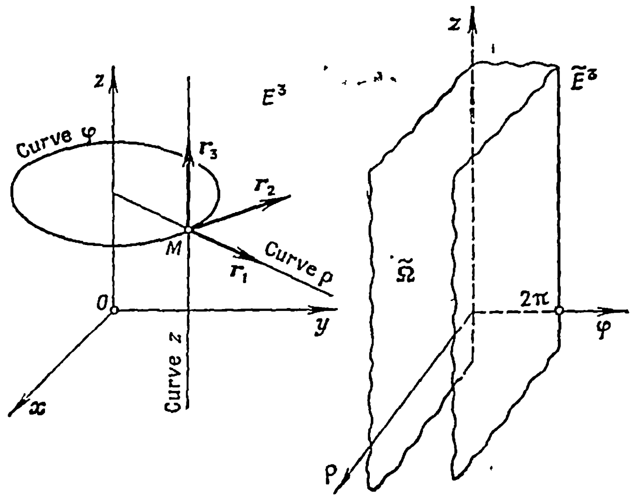


Fig. 6.3

We may therefore regard the introduction of cylindrical coordinates in E^3 as a result of mapping a domain $\tilde{\Omega}$ of $\tilde{E}^3$ into E^3 with the aid of formulas (6.60).

Obviously coordinate curves ρ (or curves x^1) are straight lines passing through the Oz axis perpendicularly to it, coordinate curves φ (curves x^2) are circles with centres on the Oz axis, whose planes are parallel to the Oxy plane. Coordinate curves z (curves x^3) are straight lines parallel to the Oz axis (see Fig. 6.3). We find the vectors $\boldsymbol{r}_1$,

r_2, r_3 and r^1, r^2, r^3. We have

$$r_1=\left\{\frac{\partial x}{\partial \rho},\ \frac{\partial y}{\partial \rho},\ \frac{\partial z}{\partial \rho}\right\}=\{\cos\varphi,\ \sin\varphi,\ 0\},$$

$$r_2=\left\{\frac{\partial x}{\partial \varphi},\ \frac{\partial y}{\partial \varphi},\ \frac{\partial z}{\partial \varphi}\right\}=\{-\rho\sin\varphi,\ \rho\cos\varphi,\ 0\},$$

$$r_3=\left\{\frac{\partial x}{\partial z},\ \frac{\partial y}{\partial z},\ \frac{\partial z}{\partial z}\right\}=\{0,\ 0,\ 1\}.$$

We stress that the expressions in the braces are the Cartesian coordinates of the basis vectors r_1, r_2, and r_3. One can directly see that the basis r_1, r_2, r_3 is orthogonal. To calculate the conjugate basis we use the formulas of Section 6.1.1. We have

$$r_1r_2r_3=\begin{vmatrix}\cos\varphi & \sin\varphi & 0\\ -\rho\sin\varphi & \rho\cos\varphi & 0\\ 0 & 0 & 1\end{vmatrix}=\rho,$$

$$[r_2r_3]=\{\rho\cos\varphi,\ \rho\sin\varphi,\ 0\},$$

$$[r_3r_1]=\{-\sin\varphi,\ \cos\varphi,\ 0\},$$

$$[r_1r_2]=\{0,\ 0,\ \rho\}.$$

Therefore

$$r^1=\frac{[r_2r_3]}{r_1r_2r_3}=\{\cos\varphi,\ \sin\varphi,\ 0\},$$

$$r^2=\frac{[r_3r_1]}{r_1r_2r_3}=\left\{-\frac{1}{\rho}\sin\varphi,\ \frac{1}{\rho}\cos\varphi,\ 0\right\},$$

$$r^3=\frac{[r_1r_2]}{r_1r_2r_3}=\{0,\ 0,\ 1\}.$$

2°. *Spherical coordinate system.* It is introduced with the aid of the relations

$$x=\rho\sin\theta\cos\varphi,\ y=\rho\sin\theta\sin\varphi,\ z=\rho\cos\theta. \tag{6.61}$$

Thus $x^1=\rho$, $x^2=\varphi$, $x^3=\theta$. It is known that the coordinates ρ, φ, θ (or, equivalently, x^1, x^2, x^3) vary within the following limits:

$$0\leqslant\rho<+\infty,\ 0\leqslant\varphi<2\pi,\ 0\leqslant\theta\leqslant\pi. \tag{6.62}$$

Inequalities (6.62) define in $\widetilde{E}^3$ with coordinates ρ, φ, θ (or x^1, x^2, x^3) an infinite domain $\widetilde{\Omega}$ represented in Fig. 6.4. We may therefore regard the introduction of spherical coordinates in E^3 as a result of mapping the domain $\widetilde{\Omega}$ of $\widetilde{E}^3$ into E^3 by means of formulas (6.61).

Obviously, coordinate curves ρ (curves x^1) are rays emanating from the origin, coordinate curves φ (curves x^2) are circles with

centres on the Oz axis, whose planes are parallel to the Oxy plane, coordinate curves θ (curves x^3) are semicircles whose centres are at the origin and whose planes pass through the Oz axis (see Fig. 6.4).

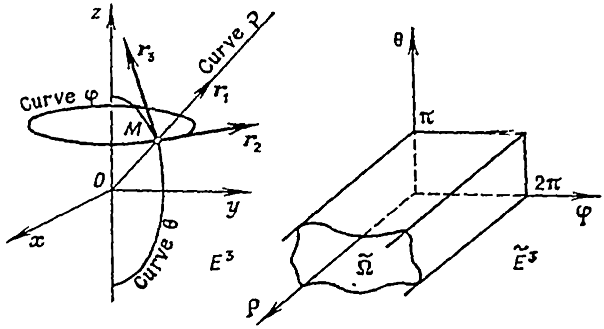


Fig. 6.4

We find the vectors r_1, r_2, r_3 and r^1, r^2, r^3. We have

$$r_1 = \{\sin\theta\cos\varphi, \quad \sin\theta, \ \sin\varphi, \quad \cos\theta\},$$
$$r_2 = \{-\rho\sin\theta\sin\varphi, \quad \rho\sin\theta\cos\varphi, \quad 0\},$$
$$r_3 = \{\rho\cos\theta\cos\varphi, \quad \rho\cos\theta\sin\varphi, \quad -\rho\sin\theta\}.$$

One can directly see that the basis r_1, r_2, r_3 is orthogonal. To calculate the conjugate basis we use the formulas of Section 6.1.1. We have

$$r_1 r_2 r_3 = \begin{vmatrix} \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta \\ -\rho\sin\theta\sin\varphi & \rho\sin\theta\cos\varphi & 0 \\ \rho\cos\theta\cos\varphi & \rho\cos\theta\sin\varphi & -\rho\sin\theta \end{vmatrix} = -\rho^2\sin\theta,$$

$$[r_2 r_3] = \{-\rho^2\sin^2\theta\cos\varphi \quad -\rho^2\sin^2\theta\sin\varphi, \quad -\rho^2\sin\theta\cos\theta\},$$
$$[r_3 r_1] = \{\rho\sin\varphi, \quad \rho\cos\varphi, \quad 0\},$$
$$[r_1 r_2] = \{-\rho\cos\theta\sin\theta\cos\varphi, \quad -\rho\cos\theta\sin\theta\sin\varphi, \quad \rho\sin^2\theta\}.$$

Therefore

$$r^1 = \frac{[r_2 r_3]}{r_1 r_2 r_3} = \{\sin\theta\cos\varphi, \sin\theta\sin\varphi, \ \cos\theta\},$$
$$r^2 = \frac{[r_3 r_1]}{r_1 r_2 r_3} = \left\{-\frac{1}{\rho}\frac{\sin\varphi}{\sin\theta}, \ -\frac{1}{\rho}\frac{\cos\varphi}{\sin\theta}, \ 0\right\},$$
$$r^3 = \frac{[r_1 r_2]}{r_1 r_2 r_3} = \left\{\frac{1}{\rho}\cos\theta\cos\varphi, \ \frac{1}{\rho}\cos\theta\sin\varphi, \ -\frac{1}{\rho}\sin\theta\right\}.$$

3°. *Orthogonal curvilinear coordinate system.* A curvilinear coordinate system is said to be *orthogonal* if so is the basis r_i defined by

equation (6.59) at each point of Ω. The cylindrical and spherical coordinate systems just considered are examples of orthogonal curvilinear coordinates.

We obtain the expression for vectors r^i of the conjugate basis for the case of the orthogonal coordinate system.

We introduce the following notation:

$$H_1 = |r_1|,\ H_2 = |r_2|,\ H_3 = |r_3|.$$

It is usual to call H_1, H_2, H_3, *Lamé* coefficients* or *parameters*.

Since the coordinate system is orthogonal and the triple of vectors r_1, r_2, r_3 is right-handed, we have

$$r_1r_2r_3 = H_1H_2H_3,\quad [r_2r_3] = \frac{H_2H_3}{H_1} r_1,\quad [r_3r_1] = \\ = \frac{H_3H_1}{H_2} r_2,\quad [r_1r_2] = \frac{H_1H_2}{H_3} r_3.$$

Using these relations and the formulas expressing the vectors of the conjugate basis in terms of vectors r_i (see Section 6.1.1) we get

$$r^1 = \frac{1}{H_1^2} r_1,\quad r^2 = \frac{1}{H_2^2} r_2,\quad r^3 = \frac{1}{H_3^2} r_3.$$

6.3.2. Expressing the gradient and directional derivative for a scalar field in curvilinear coordinates. Let $u(M)$ be a differentiable scalar field in a domain Ω in which curvilinear coordinates x^1, x^2, x^3 are introduced. Under these conditions grad u is defined at each point of Ω and at each point Ω one can calculate for any direction e a derivative $\frac{\partial u}{\partial e}$. Both the gradient grad u and the directional derivative at a given point M will be referred to the basis r_i, r^i at that point, the construction of which has been described in Section 6.3.1.

1°. *Expressing the gradient of a scalar field in curvilinear coordinates.* After introducing in Ω curvilinear coordinates x^1, x^2, x^3 the scalar field u will clearly be a function of the variables x^1, x^2, x^3:

$$u = u(x^1, x^2, x^3).$$

This function may be regarded as a result of superposing the function $u(x, y, z)$ of variables x, y, z and the functions (6.58). To calculate the derivatives $\frac{\partial u}{\partial x^i}$ therefore we may apply the indirect differentiation rule. Denoting $\frac{\partial u}{\partial x^i}$ by u_i we get

$$u_i = \frac{\partial u}{\partial x}\frac{\partial x}{\partial x^i} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial x^i} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial x^i} \tag{6.63}$$

---

* Gabriel Lamé (1795-1870) is a French mathematician.

Since $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial u}{\partial z}$ are the coordinates of the vector grad u in the basis $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ referred to the $Oxyz$ system, and $\frac{\partial x}{\partial x^i}$, $\frac{\partial y}{\partial x^i}$, $\frac{\partial z}{\partial x^i}$ are the coordinates of a vector $\boldsymbol{r}_i$, relation (6.63) may obviously be rewritten in the following form:

$$u_i = \boldsymbol{r}_i \operatorname{grad} u. \tag{6.64}$$

Using the Gibbs formulas (see formulas (6.6) of this chapter) for grad u and formulas (6.64) we get

$$\operatorname{grad} u = (\boldsymbol{r}_i \operatorname{grad} u)\, \boldsymbol{r}^i = u_i \boldsymbol{r}^i.$$

So the gradient of a scalar field u is expressed in curvilinear coordinates as

$$\operatorname{grad} u = u_i \boldsymbol{r}^i \left(u_i = \frac{\partial u}{\partial x^i} \right) \tag{6.65}$$

In practice we often encounter the case of the orthogonal curvilinear system. In Section 6.3.1 we obtained (see Section 6.3.1.3°) an expression for the vectors $\boldsymbol{r}^i$ of the conjugate basis for the orthogonal system. Using these expressions and formula (6.65) we find the following formula for grad u in orthogonal coordinates:

$$\operatorname{grad} u = \frac{1}{H_1^2} \frac{\partial u}{\partial x^1} \boldsymbol{r}_1 + \frac{1}{H_2^2} \frac{\partial u}{\partial x^2} \boldsymbol{r}_2 + \frac{1}{H_3^2} \frac{\partial u}{\partial x^3} \boldsymbol{r}_3. \tag{6.66}$$

An orthonormal basis $\boldsymbol{e}_i = \boldsymbol{r}_i / H_i$ is considered along with the orthogonal basis $\boldsymbol{r}_i$. It is easy to see that in $\boldsymbol{e}_i$ the expression for grad u is of the form

$$\operatorname{grad} u = \frac{1}{H_1} \frac{\partial u}{\partial x^1} \boldsymbol{e}_1 + \frac{1}{H_2} \frac{\partial u}{\partial x^2} \boldsymbol{e}_2 + \frac{1}{H_3^2} \frac{\partial u}{\partial x^3} \boldsymbol{e}_3. \tag{6.67}$$

2°. *Expressing the derivative of a scalar field u (M) with respect to a direction $\boldsymbol{e}$ in curvilinear coordinates.* Let e^i be the contravariant coordinates of a unit vector $\boldsymbol{e}$ in a basis $\boldsymbol{r}_i$, so that

$$\boldsymbol{e} = e^k \boldsymbol{r}_k.$$

In Section 6.2.2 we obtained the following formula for the derivative $\frac{\partial u}{\partial e}$:

$$\frac{\partial u}{\partial e} = \boldsymbol{e} \cdot \operatorname{grad} u$$

(see formula (6.40)). Substituting in this formula the expression for $\boldsymbol{e}$ in the basis $\boldsymbol{r}_i$ and formula (6.65) for grad u we get

$$\frac{\partial u}{\partial e} = (e^k \boldsymbol{r}_k)(u_i \boldsymbol{r}^i) = e^k u_i (\boldsymbol{r}_k \boldsymbol{r}^i) = e^k u_i \delta_k^i = u_i e^i.$$

Thus the derivative of a scalar field u with respect to e can be expressed in curvilinear coordinates as follows:

$$\frac{\partial u}{\partial e} = u_i e^i. \tag{6.68}$$

6.3.3. Expressing divergence, curl and directional derivative for a vector field in curvilinear coordinates. Let $p(M)$ be a differentiable vector field in a domain Ω in which curvilinear coordinates are introduced. Under these conditions the divergence and curl of the field p are defined at each point of Ω and at each point of Ω one can calculate for any direction e a derivative $\frac{\partial p}{\partial e}$. The divergence, curl and directional derivative at a given point M will be referred to the basis r_i, r^i at that point.

1°. *Expressing the divergence of a vector field in curvilinear coordinates.* After cuvilinear coordinates x^1, x^2, x^3 are introduced in Ω the vector field p will obviously be a function of the variables x^1, x^2, x^3:

$$p = p(x^1, x^2, x^3).$$

This function may be regarded as a result of superposing the function $p(x, y, z)$ and functions (6.58). To calculate the derivatives $\frac{\partial p}{\partial x^i}$ therefore, we may apply the indirect differentiation rule. Denoting $\frac{\partial p}{\partial x^i}$ by p_i we get

$$p_i = \frac{\partial p}{\partial x}\frac{\partial x}{\partial x^i} + \frac{\partial p}{\partial y}\frac{\partial y}{\partial x^i} + \frac{\partial p}{\partial z}\frac{\partial z}{\partial x^i}. \tag{6.69}$$

Since $\frac{\partial p}{\partial x} = Ai$, $\frac{\partial p}{\partial y} = Aj$, $\frac{\partial p}{\partial z} = Ak$, where A is a linear operator defined by the equation $\Delta p = A\Delta r + o(|\Delta r|)$ (see Section 6.2.3), from relations (6.69) and the properties of a linear operator we get

$$p_i = A\left(\frac{\partial x}{\partial x^i} i + \frac{\partial y}{\partial x^i} j + \frac{\partial z}{\partial x^i} k\right) = Ar_i. \tag{6.70}$$

By definition $\operatorname{div} p = \operatorname{div} A = r^i A r_i$. According to formula (6.70) therefore, in curvilinear coordinates the divergence of the vector field $p(M)$ can be calculated from the formula

$$\operatorname{div} p = r^i p_i \qquad \left(p_i = \frac{\partial p}{\partial x^i}\right). \tag{6.71}$$

We find the expression for divergence for the case of the orthogonal curvilinear coordinate system. Using the expression for the vectors r^i of the conjugate basis for orthogonal curvilinear coordinates and

formula (6.71) we get

$$\operatorname{div} p = \frac{1}{H_1^2} p_1 r_1 + \frac{1}{H_2^2} p_2 r_2 + \frac{1}{H_3^2} p_3 r_3 \quad \left(p_i = \frac{\partial p}{\partial x^i} \right). \tag{6.72}$$

There is another way of writing formula (6.72). Denote by P^i the coordinates of the field p in the orthonormal basis $e_i = \frac{r_i^*}{H_i}$. Then, after a number of transformations, the expression (6.72) for div p takes the following form:

$$\operatorname{div} p = \frac{1}{H_1 H_2 H_3} \left[\frac{\partial (P^1 H_2 H_3)}{\partial x^1} + \frac{\partial (P^2 H_3 H_1)}{\partial x^2} + \frac{\partial (P^3 H_1 H_2)}{\partial x^3} \right]. \tag{6.73}$$

2°. *Expressing the curl of a vector field in curvilinear coordinates.* By definition curl p = curl A = $[r^i A r_i]$. According to formula (6.70) therefore we get

$$\operatorname{curl} p = [r^i p_i] \quad \left(p_i = \frac{\partial p}{\partial x^i} \right). \tag{6.74}$$

We seek the expression for the curl in the orthogonal curvilinear coordinate system. Using the expression for the vectors r^i of the conjugate basis for the orthogonal system and formula (6.74) we get

$$\operatorname{curl} p = \frac{1}{H_1^2} [r_1 p_1] + \frac{1}{H_2^2} [r_2 p_2] + \frac{1}{H_3^2} [r_3 p_3] \quad \left(p_i = \frac{\partial p}{\partial x^i} \right). \tag{6.75}$$

In the orthonormal basis $e_i = \frac{r_i}{H_i}$ the curl of p has the coordinates

$$\left\{ \frac{1}{H_2 H_3} \left[\frac{\partial (P^3 H_3)}{\partial x^2} - \frac{\partial (P^2 H_2)}{\partial x^3} \right], \quad \frac{1}{H_3 H_1} \left[\frac{\partial (P^1 H_1)}{\partial x^3} - \frac{\partial (P^3 H_3)}{\partial x^1} \right], \right.$$
$$\left. \frac{1}{H_1 H_2} \left[\frac{\partial (P^2 H_2)}{\partial x^1} - \frac{\partial (P^1 H_1)}{\partial x^2} \right] \right\}. \tag{6.76}$$

3°. *Expressing the directional derivative of a vector field in curvilinear coordinates.* We use the formula

$$\frac{\partial p}{\partial e} = Ae \tag{6.46}$$

obtained in Section 6.2.3. Let $e = e^i r_i$. Then from formula (6.46) and the properties of a linear operator we get

$$\frac{\partial p}{\partial e} = e^i A r_i.$$

Since $Ar_i = p_i$, where $p_i = \frac{\partial p}{\partial x^i}$, for the derivative of the vector field p with respect to a direction e we obtain the following

* On the right-hand side of this formula we do not sum over i.

expression:

$$\frac{\partial p}{\partial c}=c^i p_i. \qquad (6.77)$$

6.3.4. Expressing the Laplacian operator in curvilinear orthogonal coordinates. We have defined the Laplacian Δu as a repeated operation div grad u. Using the expressions (6.67) and (6.73) for the gradient and divergence in curvilinear orthogonal coordinates we obtain the expression for the Laplacian.

In the case under consideration a vector field p whose divergence is to be calculated is the field grad u. Substituting (6.67) into (6.73) we get

$$\Delta u=\frac{1}{H_1H_2H_3}\left[\frac{\partial}{\partial x^1}\left(\frac{H_2H_3}{H_1}\frac{\partial u}{\partial x^1}\right)+\right.$$
$$\left.+\frac{\partial}{\partial x^2}\left(\frac{H_3H_1}{H_2}\frac{\partial u}{\partial x^2}\right)+\frac{\partial}{\partial x^3}\left(\frac{H_1H_2}{H_3}\frac{\partial u}{\partial x^3}\right)\right]. \qquad (6.78)$$

6.3.5. Expressing the basic field theory operations in cylindrical and spherical coordinate systems.

1°. *Cylindrical coordinate system.* By virtue of the results of Section 6.3.1.1° Lamé parameters for cylindrical coordinates are of the form

$$H_1=1,\quad H_2=\rho,\quad H_3=1.$$

In such a case formulas (6.67), (6.73), (6.76), and (6.78) yield the following equations:

$$\operatorname{grad} u=\frac{\partial u}{\partial \rho}e_\rho+\frac{1}{\rho}\frac{\partial u}{\partial \varphi}e_\varphi+\frac{\partial u}{\partial z}e_z,$$

$$\operatorname{div} p=\frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho P_\rho)+\frac{1}{\rho}\frac{\partial P_\varphi}{\partial \varphi}+\frac{\partial P_z}{\partial z},$$

$$\operatorname{curl} p=\left(\frac{1}{\rho}\frac{\partial P_z}{\partial \varphi}-\frac{\partial P_\varphi}{\partial z}\right)e_\rho+\left(\frac{\partial P_\rho}{\partial z}-\frac{\partial P_z}{\partial \rho}\right)e_\varphi+$$
$$+\left(\frac{1}{\rho}\frac{\partial(\rho P_\varphi)}{\partial \rho}-\frac{1}{\rho}\frac{\partial P_\rho}{\partial \varphi}\right)e_z,$$

$$\Delta u=\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial u}{\partial \rho}\right)+\frac{1}{\rho^2}\frac{\partial^2 u}{\partial \varphi^2}+\frac{\partial^2 u}{\partial z^2}.$$

2°. *Spherical coordinate system.* In this case Lamé parameters are of the form

$$H_1=1,\quad H_2=\rho\sin\theta,\quad H_3=\rho.$$

Therefore

$$\operatorname{grad} u = \frac{\partial u}{\partial \rho} e_\rho + \frac{1}{\rho \sin\theta} \frac{\partial u}{\partial \varphi} e_\varphi + \frac{1}{\rho} \frac{\partial u}{\partial \theta} e_\theta,$$

$$\operatorname{div} p = \frac{1}{\rho^2} \frac{\partial}{\partial \rho} (\rho^2 P_\rho) + \frac{1}{\rho \sin\theta} \frac{\partial P_\varphi}{\partial \varphi} + \frac{1}{\rho \sin\theta} \frac{\partial}{\partial \theta} (\sin\theta P_\theta),$$

$$\operatorname{curl} p = \frac{1}{\rho \sin\theta} \left(\frac{\partial (\sin\theta P_\varphi)}{\partial \theta} - \frac{\partial P_\theta}{\partial \varphi} \right) e_\rho +$$

$$+ \left(\frac{1}{\rho \sin\theta} \frac{\partial P_\rho}{\partial \varphi} - \frac{1}{\rho} \frac{\partial (\rho P_\varphi)}{\partial \rho} \right) e_\theta + \left(\frac{1}{\rho} \frac{\partial (\rho P_\theta)}{\partial \rho} - \frac{1}{\rho} \frac{\partial P_\rho}{\partial \theta} \right) e_\varphi,$$

$$\Delta u = \frac{1}{\rho^2} \frac{\partial}{\partial \rho} \left(\rho^2 \frac{\partial u}{\partial \rho} \right) + \frac{1}{\rho^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial u}{\partial \theta} \right) + \frac{1}{\rho^2 \sin^2\theta} \frac{\partial^2 u}{\partial \varphi^2}.$$

In conclusion we give a summary of formulas relating the operations of taking the gradient, divergence and curl to algebraic operations:

1°. $\operatorname{grad}(u \pm v) = \operatorname{grad} u \pm \operatorname{grad} v$.

2°. $\operatorname{grad}(u \cdot v) = u \operatorname{grad} v + v \operatorname{grad} u$.

3°. $\operatorname{grad}\left(\frac{u}{v}\right) = \frac{v \operatorname{grad} u - u \operatorname{grad} v}{v^2} \quad (v \neq 0)$.

4°. $\operatorname{div}(p \pm q) = \operatorname{div} p \pm \operatorname{div} q$.

5°. $\operatorname{div}(up) = p \operatorname{grad} u + u \operatorname{div} p$.

6°. $\operatorname{div}[pq] = q \operatorname{curl} p - p \operatorname{curl} q$.

7°. $\operatorname{curl}[p \pm q] = \operatorname{curl} p \pm \operatorname{curl} q$.

8°. $\operatorname{curl}(up) = u \operatorname{curl} p - [p \operatorname{grad} u]$.

It is easy for the reader to show the validity of these formulas on his own.

Concluding remarks. In this chapter we have discussed the basic operations of field theory. We have not relied on any physical ideas since our aim was to construct a mathematical theory. In the next chapter we shall derive a number of important integral relations connecting some of the operations of field theory. These relations will allow us to give physical interpretation of the concepts and operations we have introduced in the present chapter.

CHAPTER 7

THE FORMULAS OF GREEN, STOKES, AND OSTROGRADSKY

In this chapter we obtain formulas playing an important role in various applications, and in particular in field theory. They are in a sense extensions to the multidimensional case of the Newton-Liebniz formula for one-dimensional integrals.

7.1. THE GREEN* FORMULA

7.1.1. Statement of the main theorem. Let D be a finite, multiply connected in general, domain in the Oxy plane with piecewise smooth boundary L**. The domain D with the boundary L adjoined will be designated $\bar{D}$. The following *main* theorem is true.

*Theorem 7.1. Let functions $P(x, y)$ and $Q(x, y)$ be continuous in $\bar{D}$ and have continuous partial derivatives of the first order in D. If there are improper integrals of each of the partial derivatives of the functions $P(x, y)$ and $Q(x, y)$ over the domain D***, we have*

$$\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\,dy = \oint_L P\,dx + Q\,dy, \tag{7.1}$$

a relation called the **Green formula**. *The integral on the right of (7.1) is the sum of the integrals along the components of L on which a sense of rotation is indicated such that the domain D remains on the left.*

We shall first prove the Green formula for a special but sufficiently wide class of domains. We shall then establish a number of auxiliary statements which we shall need to prove the theorem we have stated.

7.1.2. The proof of the Green formula for a special class of domains. Let D be a singly connected finite domain with piecewise smooth

* George Green (1793-1841) is an English mathematician.

** L is said to be piecewise smooth if it is made up of a finite number of smooth curves. If L consists of a finite number of closed piecewise smooth curves L_i, then the connected domain D is usually said to be *multiply connected*, and L_i are said to be *connected components* of the boundary L.

*** Since the partial derivatives of $P(x, y)$ and $Q(x, y)$ exist only in an open domain D, the integrals are improper. Under the additional assumption that the partial derivatives in $\bar{D}$ are continuous the integrals become proper integrals.

boundary L. We assume that every straight line parallel to any coordinate axis intersects L in at most two points. Such domains will be called *type-K domains.*

Under the hypothesis there are improper integrals of partial derivatives of $P(x, y)$ and $Q(x, y)$. This means that for any system

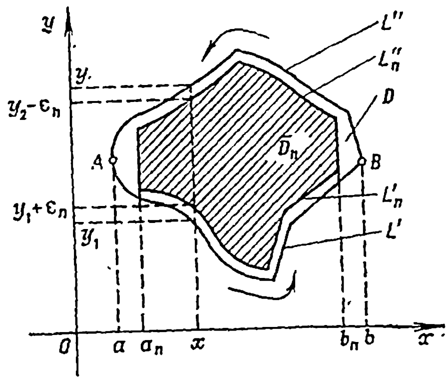


Fig. 7.1

of domains $\{D_n\}$ monotonically exhausting D we have for example

$$\lim_{n\to\infty} \iint_{\bar{D}_n} \frac{\partial Q}{\partial x}\, dx\, dy = \iint_{D} \frac{\partial Q}{\partial x}\, dx\, dy$$

(similar relations are true also for other partial derivatives of $P(x, y)$ and $Q(x, y)$).

We describe how to construct a special system of domains $\{D_n\}$ monotonically exhausting a type-K domain. We shall need it in proving the Green formula for those domains.

Let a closed interval $[a, b]$ on the Ox axis be the projection of $\bar{D}$ onto that axis. Draw through the points a and b straight lines parallel to the Oy axis. Either of the two lines intersects the boundary L in one point only. The two points A and B of intersection of the lines with L (Fig. 7.1) divide L into two curves L' and L'' which are obviously the graphs of the continuous functions $y_1(x)$ and $y_2(x)$ piecewise differentiable on $[a, b]$, respectively. Note (Fig. 7.1) that $y_1(x) \leqslant y_2(x)$ (equality holding only for $x = a$ and $x = b$).

Next consider a sequence of closed intervals $[a_n, b_n]$ such that $a < a_n < b_n < b$, $a_n \to a$, $b_n \to b$ as $n \to \infty$. In addition suppose that for any n an interval $[a_n, b_n]$ is contained in $[a_{n+1}, b_{n+1}]$. Choose a number $\varepsilon_n > 0$ so that the graphs L'_n and L''_n of the functions $y_1(x) + \varepsilon_n$ and $y_2(x) - \varepsilon_n$ are in D and do not intersect.

The boundary of $\overline{D}_n$ is a curve made up of the curves L'_n and L''_n and the segments of the vertical straight lines passing through the points a_n and b_n (Fig. 7.1). The domain $\overline{D}_{n+1}$ is constructed in a similar way, but instead of the interval $[a_n, b_n]$ we take an interval $[a_{n+1}, b_{n+1}]$ and choose a number $\varepsilon_{n+1} > 0$ less than the number ε_n. It is obvious that if $\varepsilon_n \to 0$, then the constructed system of domains $\{\overline{D}_n\}$ monotonically exhausts the domain D.

We prove the following *statement*.

Theorem 7.2. Let functions $P(x, y)$ and $Q(x, y)$ satisfy in a type-K domain D the hypotheses of Theorem 7.1. Then for that domain and for the functions $P(x, y)$ and $Q(x, y)$ the Green formula is valid.

Proof. It suffices to show the validity of the equations

$$\iint_D \frac{\partial Q}{\partial x}\,dx\,dy = \oint_L Q\,dy, \quad -\iint_D \frac{\partial P}{\partial y}\,dx\,dy = \oint_L P\,dx. \tag{7.2}$$

Since the proofs for these are of the same type, we shall prove only the second of the equations.

Consider the double integral

$$\iint_{\overline{D}_n} \frac{\partial P}{\partial y}\,dx\,dy. \tag{7.3}$$

For $\overline{D}_n$ and the integrand $\frac{\partial P}{\partial y}$ in (7.3) all the conditions hold under which the repeated integration formula is valid. From this formula we have

$$\iint_{\overline{D}_n} \frac{\partial P}{\partial y}\,dx\,dy = \int_{a_n}^{b_n} dx \int_{y_1(x)+\varepsilon_n}^{y_2(x)-\varepsilon_n} \frac{\partial P}{\partial y}\,dy =$$

$$= \int_{a_n}^{b_n} P(x, y_2(x) - \varepsilon_n)\,dx - \int_{a_n}^{b_n} P(x, y_1(x) + \varepsilon_n)\,dx. \tag{7.4}$$

The left-hand side of relations (7.4) as $n \to \infty$ has a limit equal to $\iint_D \frac{\partial P}{\partial y}\,dx\,dy$. By uniform continuity of $P(x, y)$ in $\overline{D}$, each of the terms on the right of (7.4) has as $n \to \infty$ a limit equal to $\int_a^b P(x, y_2(x))\,dx$ for the first term and to $\int_a^b P(x, y_1(x))\,dx$ for the second. The former integral, with the direction of circulation about the boundary as indicated in Fig. 7.1, is the line integral

$$-\int_{L''} P(x, y)\,dx$$

ant nature, its form and value remain unaltered under a transition to a new Cartesian system. Indeed, under such a transformation of coordinates the absolute value of the Jacobian transformation is equal to unity. According to the remark, however, the integrand changes neither form nor value.

Now consider the integral

$$\oint_L P\,dx + Q\,dy \tag{7.7}$$

on the right of the Green formula. We show that *this integral has also an invariant nature, its form and value remaining unaltered under a transition to a new Cartesian system.*

Let t be the unit tangent vector at the points of L whose direction agrees with the direction of circulation about L, and let $\cos\alpha$ and $\sin\alpha$ be the coordinates of t. Choose as a parameter on L an arc length l, the increase in l agreeing on each connected component of the boundary with the direction of circulation about that component. Under the conditions on L the function $t\,(l)$ will be piecewise continuous. Under the conditions formulated above the vector field p is continuous on L and its coordinates P and Q are continuous functions of l.

Notice that once the sense of rotation and the parameter of L are chosen the line integral of the second kind (7.7) is transformed into a line integral of the first kind. P and Q are calculated at the points of L and $dx = \cos\alpha\,dl$, $dy = \sin\alpha\,dl$. Thus

$$\oint_L P\,dx + Q\,dy = \oint_L (P\cos\alpha + Q\sin\alpha)\,dl = \oint_L pt\,dl. \tag{7.8}$$

Relation (7.8) shows that the integral (7.7) has indeed an invariant nature: the scalar product pt is an invariant, and parametrization with the aid of arc length is independent of the coordinate system. Moreover, in the new Cartesian system $Ox'y'$

$$pt\,dl = (P'\cos\alpha' + Q'\sin\alpha')\,dl = P'\,dx' + Q'\,dy',$$

and therefore

$$P\,dx + Q\,dy = P'\,dx' + Q'\,dy'.$$

So we have shown that the integral (7.7) has an invariant nature, its form and value remain unaltered under a transition to a new Cartesian system.

The above reasoning allows the Green formula (7.1) to assume the following *invariant form*

$$\iint_D k\operatorname{curl} p\,d\sigma = \oint_L pt\,dl, \tag{7.9}$$

where $d\sigma$ is an area element of D.

Remark 2. The integral

$$\oint_L pt\, dl$$

is usually called the *circulation of a vector field p about a curve L.*

From Theorem 7.2 and the conclusions of this subsection we may draw an important corollary.

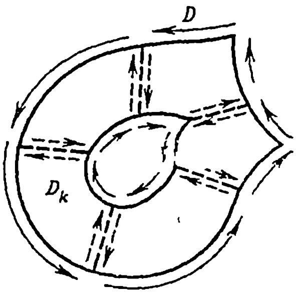


Fig. 7.2

Corollary. Let functions $P(x, y)$ and $Q(x, y)$ satisfy the hypotheses of Theorem 7.1 in a finite domain D with piecewise smooth boundary L. If D can be divided into a finite number of domains D_k with piecewise smooth boundaries L_k (Fig. 7.2) *each D_k being a type-K domain with respect to some Cartesian system, then for D and for $P(x, y)$ and $Q(x, y)$ the Green formula is valid.*

The validity of the corollary follows from the following arguments. It is clear that the Green formula is valid for each of the domains D_k. This follows from the invariance of the formula and from Theorem 7.2 (in some coordinate system D_k is a type-K domain).

It is also obvious that the sum of the integrals $\iint_{D_k} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy$ on the left-hand sides of the Green formulas over the domains D_k is $\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy$. But the sum of the line integrals $\oint_{L_k} P\, dx + Q\, dy$ on the right-hand sides of the Green formulas along the boundaries L_k of D_k gives $\oint_L P\, dx + Q\, dy$, for the integrals along the common parts of the boundary of the domains D_k cancel out, these parts being traced in opposite directions in adjacent domains D_k (consider Fig. 7.2 for explanation).

Remark 3. An arbitrary finite connected domain D with piecewise smooth boundary L cannot in general be divided into a finite number of domains D_k of the type indicated above. From every finite domain D with piecewise smooth boundary, however, we can remove an arbitrarily small part such that the remaining domain can be divided in the necessary way. Contributions to the right- and left-hand sides of the Green formula corresponding to the removed part of D will accordingly be arbitrarily small. This idea is the basis of the proof for the Green formula in the general case.

In Section 7.1.4 we shall prove a number of auxiliary propositions using which we shall establish the Green formula in the general case in the way indicated above.

7.1.4. **Auxiliary propositions.** Let L be a piecewise smooth plane curve without self-intersections on which the arc length l is chosen as the parameter.

A neighbourhood of an interior point P on L is any connected open set of points of L not coinciding with the whole of the curve L and containing the point P. *For an end point of* L we introduce the concept of *half-neighbourhood.** The length of a neighbourhood (or half-neighbourhood) is called its extent.

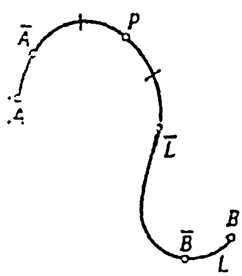


Fig. 7.3

An interior point P of L divides each of its neighbourhoods into two half-neighbourhoods. A neighbourhood of P is said to be a λ-neighbourhood if either of the half-neighbourhoods is of length λ.

Lemma 1. Let L be a smooth finite curve without self-intersections, let A and B be the end points of the curve, and let $\bar{L}$ be a connected part of L consisting together with its end points $\bar{A}$ and $\bar{B}$ entirely of interior points of L (Fig. 7.3)**. *We can find two positive numbers λ and δ such that the supremum of the angles which the tangents at the points of the λ-neighbourhood of any point P of $\bar{L}$*** make with the tangent at P is less than $\pi/8$ and the distances from P to the points of L outside the λ-neighbourhood are less than δ****.*

Proof. *We show that we can find $\lambda > 0$ satisfying the hypotheses of the lemma.* First, we note that given any $\alpha > 0$ we can find *for every* point P a λ-neighbourhood ($\lambda > 0$) such that within it the supremum of the angles the tangents at the points of that λ-neighbourhood make with the tangent at the point P is less than α. This follows from the continuity of tangents to the curve L.

The question is whether there is a universal λ suitable for *all points* of the curve $\bar{L}$.

Assume that there is no $\lambda > 0$ satisfying the hypotheses of the lemma. Then given any $\lambda_n = 1/n$ on $\bar{L}$ we can find points P_n and

---

* If P is an end point of L and Q is any of its other points, then the set of all points of L between P and Q including P but not Q is said to be *a half-neighbourhood* of P.

** The curve may be closed. In this case $\bar{L}$ may coincide with L. If L is a closed curve with one corner point, then $\bar{L}$ is any closed connected part of L not containing that corner point.

*** A neighbourhood of a point of $\bar{L}$ is considered to be a neighbourhood of that point on L.

**** Obviously $\lambda \geqslant \delta$.

Q_n such that the length of the arc P_nQ_n is less than λ_n and the angle between the tangents at those points is at least equal to a fixed $\alpha < \pi/8$. We select a subsequence $\{P_{n_k}\}$ of $\{P_n\}$ converging to a point P of $\bar{L}$. Obviously $\{Q_{n_k}\}$ also converges to P. Consider the λ-neighbourhood of P in which the supremum of the angles between the tangents at the points of the neighbourhood and at P is less than $\alpha/2$.

It is clear that the angle between the tangents at any two points of that λ-neighbourhood of P is less than α. With n_k sufficiently large, the points P_{n_k} and Q_{n_k} will be in the chosen λ-neighbourhood of P and therefore the angle between the tangents at those points must be less than α whereas by the choice of the points it must be greater or equal to α. This contradicts our assumption that there is no $\lambda > 0$ satisfying the hypotheses of the lemma. Note that the required λ is less than either of the arcs $A\bar{A}$ or $B\bar{B}$.

Now we prove that we can find $\delta > 0$ *satisfying the hypotheses of the lemma.*

Assume that there is no $\delta > 0$ satisfying the hypotheses of the lemma. Then given any $\delta_n = 1/n$ we can find a point P_n on $\bar{L}$ and a point Q_n on L such that the length of the arc P_nQ_n is equal to or greater than λ* whereas the chord P_nQ_n is of length smaller than δ_n. We select a subsequence of $\{P_n\}$ converging to a point P of $\bar{L}$ and consider the corresponding subsequence of $\{Q_n\}$. We select a subsequence $\{Q_{n_k}\}$ of that last subsequence converging to the point Q of L. It is clear that $\{P_{n_k}\}$ converges to P. Since by the choice of the points P_{n_k} and Q_{n_k} the length of the arc $P_{n_k}Q_{n_k}$ is equal to or greater than λ, so is the length of the arc PQ. Since the lengths of the chords $P_{n_k}Q_{n_k}$ tend to zero, the length of the chord PQ is zero, i.e. the point P coincides with Q and is therefore a point of self-intersection of the curve L without self-intersections. This contradiction supports the possibility of choosing the required $\delta > 0$. The proof of the lemma is complete.

Corollary 1. Let the curves L and $\bar{L}$ satisfy the hypotheses of Lemma 1. Then we can find a number 2λ such that any arc of $\bar{L}$ of length smaller than 2λ is 1-1 projected onto one of the coordinate axes of a fixed Cartesian system Oxy.

Indeed, take as λ the number indicated in Lemma 1. Any arc of $\bar{L}$ of length smaller than 2λ is contained in a λ-neighbourhood of some point P on $\bar{L}$. The tangent at P makes with the Ox or Oy axis an angle equal to or less than $\pi/4$. Then clearly the tangent at any point of the arc under consideration makes with that axis an angle *less than* $\pi/2$ and therefore the arc is 1-1 projected onto the

* That such a λ exists has been established in the first part of the proof.

axis (were it not 1-1 projected there would be tangents making with the axis an angle equal to $\pi/2$).

Corollary 2. Let the curves L and $\bar{L}$ satisfy the hypotheses of Lemma 1. Then we can find a number $2\lambda > 0$ such that any arc of $\bar{L}$ of length smaller than 2λ is 1-1 projected onto both coordinate axes of a Cartesian system Oxy especially chosen for that arc.

Take as λ the number indicated in the lemma. Any arc of $\bar{L}$ less than 2λ is contained in a λ-neighbourhood of some point P of $\bar{L}$. Choose a Cartesian system so that the tangent at P makes with the coordinate axes an angle $\pi/4$. Then the tangent at any point of the arc will make with either of the Ox and Oy axes an angle less than $\pi/2$ and therefore the arc will be 1-1 projected onto either of the axes. Note that slight changes in the coordinate system chosen do not affect the possibility of the arc being 1-1 projected onto both coordinate axes.

Lemma 2. Let Q be a square, and let R be an angle with vertex at the centre P of the square Q and with opening $2\alpha < \pi/4$. Denote by Γ the part of the boundary of Q contained in R. Then the angle between any chord of the curve Γ (the straight line joining two points of Γ) and the bisectrix of R is at least α.

In view of the elementary nature of the lemma we leave it unproved.

Lemma 3. Let Q be a square and let L be a smooth curve without self-intersections emanating from the centre P of Q. Also suppose that the supremum of the angles the tangents to L make with the half-tangent to L at P is equal to $\bar{\alpha} < \pi/8$. Then L intersects the boundary of Q in at most one point.

Proof. Construct an angle R of opening 2α, $2\bar{\alpha} < 2\alpha < \pi/4$ whose bisectrix is the half-tangent to L at P and whose vertex is the centre P of the square. Denote by Γ the part of the boundary of Q contained in R. Obviously L is inside R (if L did not intersect a side of R at a point other than P, then there would be a tangent parallel to that side and the tangent would make with the half-tangent to L at P an angle equal to $\alpha > \bar{\alpha}$, which contradicts the hypothesis). Let L intersect Γ in two points M and N. Then on L we could find a point the tangent at which would be parallel to the chord MN and according to Lemma 3 that tangent would make with the half-tangent to L at P an angle at least equal to $\alpha > \bar{\alpha}$ and this contradicts the hypothesis. The proof of the lemma is complete.

Corollary of Lemmas 1 and 3. Let L and $\bar{L}$ satisfy the hypotheses of Lemma 1 and let $\delta > 0$ be the number indicated in that lemma. Then $\bar{L}$ intersects the boundary of any square Q with centre at an arbitrary point P and with side of less that $\sqrt{2}\delta$ in at most two points.

We shall demonstrate the validity of the corollary. Let P be an arbitrary point of $\bar{L}$ and let $\lambda > 0$ be the number indicated in

Lemma 1. Consider a λ-neighbourhood of P. Both end points of the neighbourhood and the part of $\bar{L}$ outside the λ-neighbourhood are according to Lemma 1 outside any square with centre at P and of side less than $\sqrt{2}\delta$. Therefore the λ-neighbourhood under consideration (and only that neighbourhood) intersects the boundary of Q^*. Since either of the half-neighbourhoods of the λ-neighbourhood of P under consideration satisfies the hypotheses of Lemma 3, it is clear that the λ-neighbourhood intersects the boundary of Q in at most two points.

7.1.5. Special subdivision of a domain D with piecewise smooth boundary L. Let D be a multiply connected domain whose boundary L consists of a finite number of closed piecewise smooth curves, $P_1, P_2, \ldots, P_N$ being the corner points of L. We assume a Cartesian system Oxy to be chosen in the plane.

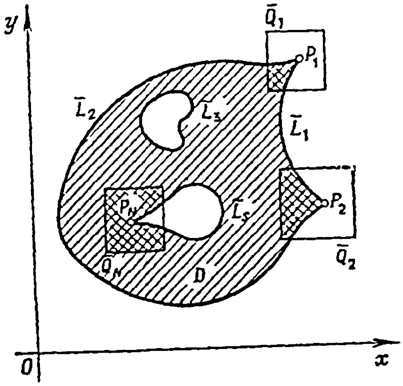


Fig. 7.4

We shall show a method of special subdivision of D into subdomains. We shall need such subdivisions in proving Theorem 7.1.

1°. We show that for any $\varepsilon > 0$ we can choose the squares $Q_1, Q_2, \ldots$ $\ldots, Q_N$ with centres at the corner points of L and with sides parallel to the Ox and Oy axes (Fig. 7.4) so that the following *conditions* hold:

(1) The boundary of any square $\bar{Q}_i$ with centre at P_i intersects either of the two branches of L emanating from P_i** in exactly one point (see Fig. 7.4). Those points are the only points the boundary of a square Q_i has in common with that of L.

(2) The sum of the areas of Q_i is less than ε; the sum of the lengths of the parts of L that are in the squares $\bar{Q}_i$ is also less than ε. The sum of the perimeters of the squares $\bar{Q}_i$ clearly is not greater than $A\varepsilon$, where A is some constant.

The possibility of the above choice of the squares $\bar{Q}_i$ arises from the following arguments.

Consider the λ-neighbourhoods of corner points subject to the requirements:

1. These λ-neighbourhoods do not intersect.

---

* Here we are using the Jordan theorem which states that if two points of a continuous curve L are an interior and an exterior point of a domain D then L intersects the boundary of D.

** A sufficiently small λ-neighbourhood of a corner point P_i consists of two smooth branches emanating from that point.

2. The sum of the lengths of all λ-neighbourhoods is less than ε.

3. The supremum of the angles the tangents of either of the half-neighbourhoods of a λ-neighbourhood made with the corresponding half-tangent at a corner point is less than $\bar{\alpha} < \pi/8$. The possibility of the choice of such λ-neighbourhoods of corner points is obvious. Note that either of the half-neighbourhoods of the λ-neighbourhoods chosen satisfies the hypotheses of Lemma 3. Therefore either of these half-neighbourhoods intersects in at most one point the boundary of any square with centre at the corresponding corner point.

For every corner point P_i we define a number $\delta_i > 0$ equal to the infimum of the distances from P_i to the part of L obtained by removing from L a λ-neighbourhood of the point P_i.

We denote $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_N\}$. It is clear that any square $\bar{Q}_i$ with centre at P_i the length of whose side is less than $\sqrt{2}\,\delta$ satisfies the above *condition* (1) for, with the choice of square $\bar{Q}_i$ indicated, the hypotheses of Lemma 4 hold for either of the half-neighbourhoods of the point P_i and, in addition, the end points of the half-neighbourhood are outside the square $\bar{Q}_i$ (this ensures the uniqueness of the point of intersection of the half-neighbourhood with the boundary of the square). It is also clear that by making smaller the sides of the squares we can make the sum of their areas become less than ε. Obviously, the sum of the lengths of the parts of L that are in the squares $\bar{Q}_i$ will be less than ε owing to a special choice of λ-neighbourhoods of the corner points. Thus *condition* (2) also holds when $\bar{Q}_i$ are chosen as indicated.

2°. We remove from L those parts which are in the squares $\bar{Q}_i$. The remaining part of L is a collection of smooth curves $\bar{L}_i$ without common points, some of $\bar{L}_i$ being smooth closed curves. Note that every open curve $\bar{L}_i$ consists of interior points of a smooth curve L_i whose end points are the corner points of L (see Fig. 7.4).

For each of the curves $\bar{L}_i$ we use Lemma 1 of the preceding subsection. Let λ_i and δ_i^* be numbers guaranteed for $\bar{L}_i$ by that lemma. We make δ_i^* obey yet another requirement, that δ_i^* should be less than the infimum of the distances from the points of $\bar{L}_i$ to the other curves $\bar{L}_k$. Further denote $\lambda = \min\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ and $0 < \delta^* < \min\{\delta_1^*, \delta_2^*, \ldots, \delta_N^*\}$, $\delta^* < \sqrt{2}\,\delta$, where δ is the number chosen in 1°. Obviously $\lambda \geqslant \delta^*$.

We divide each curve $\bar{L}_i$ into a finite number of parts of length less than δ^*. We construct squares Q_i whose centres are at division points of the curve $\bar{L}_i$, with sides of length δ^* parallel to the Ox and Oy axes.

3°. Using squares $\bar{Q}_i$ and Q_i we construct the required subdivision of the domain D.

(1) Remove from D the parts D and the squares Q_i have in common. The remaining part of D is designated $\overline{D}_\varepsilon$ and the boundary of $\overline{D}_\varepsilon$ is designated $\overline{L}_\varepsilon$. The boundary $\overline{L}_\varepsilon$ consists of curves $\overline{L}_i$ and the line segments parallel to the coordinate axes.

(2) Denote by $\tilde{Q}_i$ the part a square Q_i and the domain $\overline{D}_\varepsilon$ have in common. The domains $\tilde{Q}_i$ divide the domain $\overline{D}_\varepsilon$ into singly connected parts $\overline{D}_i^*$, the boundary of each consisting of line segments parallel to the coordinate axes and of possibly one curvilinear segment contained in one of the curves $\overline{L}_i$ and having a length smaller than δ. Since that curvilinear segment is 1-1 projected onto one of the coordinate axes (the length of each of such segments is less than $\delta^* < \lambda$ and in this case according to Corollary 1 of Lemma 1 it is 1-1 projected onto one of the coordinate axes), *any domain $\overline{D}_i$ clearly can be divided with straight lines parallel to one of the coordinate axes into a finite number of parts D_k, each a rectangle or a curvilinear trapezoid** possibly degenerated into curvilinear triangle.*

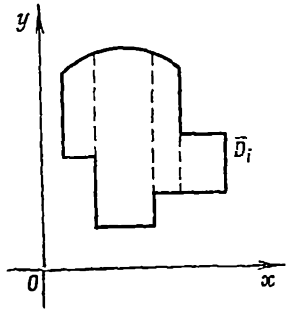


Fig. 7.5

Figure 7.5. shows a domain $\overline{D}_i$, the dotted lines indicating a subdivision of $\overline{D}_i$ into parts D_k.

7.1.6. Proof of theorem 7.1. We have just seen that after removing from D the parts that are in squares $\overline{Q}_i$ we obtain a domain $\overline{D}_\varepsilon^{***}$ with boundary L_ε which can be divided into a finite number of domains D_k of special kind.

We prove that the Green formula is true for $\overline{D}_\varepsilon$. According to the corollary of Section 7.1.3, to do this it is sufficient to show that each of D_k is a type-K domain with respect to some specially chosen Cartesian system.

If D_k is a rectangle, then the required system is for example a coordinate system one of whose axes is parallel to a diagonal of the rectangle. Let D_k be a curvilinear trapezoid or a curvilinear triangle. From the way D_k are constructed it follows that the curved side of

* A domain D is said to be singly connected if any piecewise smooth, not self-intersecting closed curve in D bounds a domain all points of which are in D.

** Recall that a curvilinear trapezoid is a figure whose bases are parallel to one of the coordinate axes, one of whose lateral sides is parallel to the other coordinate axis and the curved lateral ide is 1-1 projected onto that axis.

*** Recall that $\overline{Q}_i$ are chosen for any given positive ε so that the sum of their areas is less than ε and the sum of the lengths of the parts of the boundary L that are in $\overline{Q}_i$ is also less than ε. It is clear that as $\varepsilon \to 0$ the domains $\overline{D}_\varepsilon$ exhaust D.

the boundary of D_h satisfies the hypotheses of Lemma 1 in Section 7.1.4 and therefore, according to Corollary 2 of the lemma, is 1-1 projected onto both coordinate axes of a specially chosen Cartesian system. Since slight changes in the chosen system do not violate the indicated property, we may clearly choose a coordinate system such that the rectilinear parts of the boundary of D_h are also 1-1 projected onto its both axes. With respect to that system D_h is a type-K domain. So for D_ε the Green formula

$$\iint\limits_{D_\varepsilon} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy = \oint\limits_{L_\varepsilon} P\, dx + Q\, dy \tag{7.10}$$

is true. From the way domains D_ε are constructed it follows that as $\varepsilon \to 0$ the left- and right-hand sides of formula (7.10) have respectively the limits $\iint\limits_{D} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy$ and $\int\limits_{L} P\, dx + Q\, dy$. The proof of Theorem 7.1 is complete.

7.2. THE STOKES * FORMULA

7.2.1. Statement of the main theorem. Let S be a bounded complete piecewise smooth two-sided surface with piecewise smooth boundary Γ**.

A neighbourhood of S is any open set Ω containing S.

The following *main* theorem is true.

Theorem 7.3. Let functions $P(x, y, z)$, $Q(x, y, z)$ and $R(x, y, z)$ be continuous and have continuous partial derivatives of the first order in some neighbourhood of a surface S. Then the following relation holds

$$\iint\limits_{S} \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy\, dz +$$
$$+\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz\, dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy =$$
$$= \oint\limits_{\Gamma} P\, dx + Q\, dy + R\, dz \tag{7.11}$$

called the Stokes formula. The integral on the right-hand side is a sum of integrals along the connected components of Γ on which a direction of circulation is indicated such that with regard to the choice of surface side the surface S remains on the left.

Using Remark 2 of Section 5.3.2 on the form of notation for surface integrals of the second kind and symbols X, Y, Z for the angles

* George Gabriel Stokes (1819-1903) is a well-known English physicist and mathematician.

** Note that a closed surface has no boundary.

the normal to the surface forms with the coordinate axes we can rewrite the Stokes formula (7.11) as follows:

$$\iint_S \left[\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\cos X+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\cos Y+\right.$$
$$\left.+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\cos Z\right]d\sigma=\oint_\Gamma P\,dx+Q\,dy+R\,dz. \tag{7.12}$$

In the subsections that follow we shall prove a number of propositions we shall need to prove the stated theorem.

7.2.2. The proof of the Stokes formula for a smooth surface 1-1 projected onto three coordinate planes. The following theorem is true.

Theorem 7.4. Let S be a bounded complete smooth two-sided singly connected surface with piecewise smooth boundary Γ. We assume S to be 1-1 projected onto each of the coordinate planes of a system $Oxyz$. Let P, Q, and R be functions given in some neighbourhood of S, continuous in that neighbourhood and having in it continuous partial derivatives of the first order. Then the Stokes formula (7.11) is valid.

Proof. To prove the theorem we turn to the form (7.12) of writing the Stokes formula. We shall assume the unit normal vectors to form acute angles with the coordinate axes.

Obviously the theorem will be proved if we prove the equations

$$\left.\begin{aligned}
&\iint_S \left(\frac{\partial P}{\partial z}\cos Y-\frac{\partial P}{\partial y}\cos Z\right)d\sigma=\oint_\Gamma P\,dx,\\
&\iint_S \left(\frac{\partial Q}{\partial x}\cos Z-\frac{\partial Q}{\partial z}\cos X\right)d\sigma=\oint_\Gamma Q\,dy,\\
&\iint_R \left(\frac{\partial R}{\partial y}\cos X-\frac{\partial R}{\partial x}\cos Y\right)d\sigma=\oint_\Gamma R\,dz.
\end{aligned}\right\} \tag{7.13}$$

Since relations (7.13) yield themselves to proofs of the same type, we shall prove only the first of them.

We denote by I the integral on the left of the first of the equations (7.13):

$$I=\iint_S \left(\frac{\partial P}{\partial z}\cos Y-\frac{\partial P}{\partial y}\cos Z\right)d\sigma. \tag{7.14}$$

Under the hypothesis the surface S is smooth and 1-1 projected onto the Oxy plane. Therefore S is the graph of the differentiable function $z=z(x, y)$. In this case, with regard to the orientation of the unit normals to S, $\cos Y$ and $\cos Z$ can be found from the

formulas

$$\cos Y = \frac{-q}{\sqrt{1+p^2+q^2}}, \quad \cos Z = \frac{1}{\sqrt{1+p^2+q^2}}, \tag{7.15}$$

where $p=\frac{\partial z}{\partial x}$, $q=\frac{\partial z}{\partial y}$.

Using formulas (7.15) relation (7.14) may be rewritten as follows:

$$I = -\iint\limits_S \left(\frac{\partial P}{\partial y} + q\frac{\partial P}{\partial z}\right)\cos Z\, d\sigma. \tag{7.16}$$

Since on S the values of $P(x, y, z)$ are equal to $P(x, y, z(x, y))$, by the indirect differentiation rule

$$\frac{\partial}{\partial y}[P(x, y, z(x, y))] = \frac{\partial P}{\partial y} + q\frac{\partial P}{\partial z}.$$

Therefore relation (7.16) becomes

$$I = -\iint\limits_S \frac{\partial}{\partial y}[P(x, y, z(x, y))]\cos Z\, d\sigma. \tag{7.17}$$

Let D be the projection onto the Oxy plane of the surface S and let L be the projection onto that plane of the boundary Γ of S. Obviously the surface integral on the right of (7.17) is equal to the double integral $\iint\limits_D \frac{\partial}{\partial y}[P(x, y, z(x, y))]\,dx\,dy$ (see Remark 2 in Section 5.3.2) and therefore

$$I = -\iint\limits_D \frac{\partial}{\partial y}[P(x, y, z(x, y))]\,dx\,dy. \tag{7.18}$$

Applying to the integral on the right of (7.18) the Green formula we get

$$\iint\limits_D \frac{\partial}{\partial y}[P(x, y, z(x, y))]\,dx\,dy = -\oint\limits_L P(x, y, z(x, y))\,dx. \tag{7.19}$$

Suppose a point $M(x, y, z)$ of Γ is projected into the point $N(x, y)$ of the curve L. Then clearly the value of the function $P(x, y, z)$ at the point M of Γ coincides with the value of $P(x, y, z(x, y))$ at the point N of L. Therefore

$$\oint\limits_L P(x, y, z(x, y))\,dx = \oint\limits_\Gamma P(x, y, z)\,dx. \tag{7.20}$$

Obviously relations (7.14), (7.18) to (7.20) yield the first of the equations (7.13). The second and third of these equations are proved

similarly, but it is necessary to consider the projections of S onto the Oyz and Oxz planes respectively. The proof of the theorem is complete.

7.2.3. The invariant form of the Stokes formula. Let functions $P(x, y, z)$, $Q(x, y, z)$, and $R(x, y, z)$ be continuous and have continuous partial derivatives of the first order in some neighbourhood Ω of a surface S. We define in Ω a vector field $\boldsymbol{p}$ whose coordinates in a given Cartesian system are equal to P, Q, R. Obviously, under the conditions imposed on the functions P, Q, R the fild $\boldsymbol{p}$ is continuous and differentiable in Ω. We find the curl of $\boldsymbol{p}$. Using the expression for curl $\boldsymbol{p}$ in the orthonormal basis $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ we get

$$\operatorname{curl} \boldsymbol{p}=\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right) \boldsymbol{i}+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right) \boldsymbol{j}+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right) \boldsymbol{k}. \quad (7.21)$$

We choose on S a definite side, i.e. indicate on S a continuous field of unit normals $\boldsymbol{n}$. Turning to the expression (7.21) for curl $\boldsymbol{p}$ and using the standard notation $\cos X$, $\cos Y$, $\cos Z$ for the coordinates of the unit normal vector $\boldsymbol{n}$ to S we get

$$\boldsymbol{n} \operatorname{curl} \boldsymbol{p}=\left(\frac{\partial R}{\partial x}-\frac{\partial Q}{\partial z}\right) \cos X+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right) \cos Y+$$
$$+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right) \cos Z. \quad (7.22)$$

It follows from relation (7.22) that the integral on the left of the Stokes formula (7.12) may be written as

$$\iint_S \boldsymbol{n} \operatorname{curl} \boldsymbol{p}\, d\sigma. \quad (7.23)$$

So once a definite side of the surface is chosen the integral on the left side of formula (7.12) may be regarded as the surface integral of the first kind (7.23) of the function $\boldsymbol{n}$ curl $\boldsymbol{p}$ given on S. Since the scalar product $\boldsymbol{p}$ curl $\boldsymbol{p}$ and the area element $d\sigma$ of S are independent of the choice of Cartesian system in space, under a change to a new orthonormal basis $\boldsymbol{i}'$, $\boldsymbol{j}'$, $\boldsymbol{k}'$ the left-hand side of formula (7.12) remains unaltered in form and value, i.e. it is *invariant* under a choice of Cartesian system in space.

Now consider the integral

$$\oint_\Gamma P\, dx+Q\, dy+R\, dz \quad (7.24)$$

on the right of the Stokes formula.

We show that *this integral also has an invariant nature*, its form and value remain unaffected under a transformation to a new Cartesian system.

Let t be a unit tangent vector at the points of the boundary Γ of S whose direction agrees with the direction of circulation about Γ, and let $\cos\alpha$ $\cos\beta$, $\cos\gamma$ be the coordinates of t. Choose the arc length l to be the parameter of Γ, increases in the parameter on each connected component of Γ agreeing with the direction of circulation about that component. Under the conditions imposed on Γ the function $t(l)$ is piecewise continuous. Since p is continuous on Γ, its coordinates are continuous functions of l on Γ. Note that once the direction of circulation and the parameter on Γ are chosen the line integral of the second kind (7.24) transforms into a line integral of the first kind, with P, Q and R calculated at the points of Γ and $dx = \cos\alpha\, dl$, $dy = \cos\beta\, dl$, $dz = \cos\gamma\, dl$. Thus

$$\oint_\Gamma P\,dx + Q\,dy + R\,dz =$$

$$= \oint_\Gamma (P\cos\alpha + Q\cos\beta + R\cos\gamma)\,dl = \oint_\Gamma pt\,dl. \tag{7.25}$$

Relations (7.25) show that the integral (7.24) has indeed an invariant nature: the scalar product pt is an invariant, parametrization with the aid of the arc length is independent of the coordinate system.

In the new coordinate system $Ox'y'z'$ we have

$$pt\,dl = (P'\cos\alpha' + Q'\cos\beta' + R'\cos\gamma')\,dl = P'\,dx' +$$
$$+ Q'\,dy' + R'\,dz'.$$

Therefore

$$P\,dx + Q\,dy + R\,dz = P'\,dx' + Q'\,dy' + R'\,dz'.$$

Note that the integral

$$\oint_\Gamma pt\,dl$$

is usually called *the circulation of a vector field p about the curve Γ*.

The above reasoning allows the Stokes formula (7.11) (or (7.12)) to assume the following invariant form:

$$\iint_S n \operatorname{curl} p\, d\sigma = \oint_\Gamma pt\,dl. \tag{7.26}$$

7.2.4. **Proof of Theorem 7.3.** We prove the following auxiliary statement.

Lemma. Let S be a bounded complete two-sided smooth surface with piecewise smooth boundary Γ. There is $\delta > 0$ such that any con-*

* Note that a closed surface has no boundary.

nected part of S whose size is less than δ is 1-1 projected onto each of the coordinate planes of some Cartesian system.*

Proof. We first show that some neighbourhood of each point M of such a surface is 1-1 projected onto each of the coordinate planes of some Cartesian system.

Let n_M be a unit normal vector to the surface at M. Choose a Cartesian system $Oxyz$ so that n_M makes acute angles with the Ox, Oy, and Oz axes. Then in that coordinate system the determinants

$$\begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix} \quad \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \quad \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix},$$

are clearly nonzero for the values of u and v determining the point M, and by the smoothness of S they are nonzero in some neighbourhood of the point (u, v) (these determinants are proportional to the coordinates of the unit normal vector to the surface). Turning to the proof of Theorem 5.1 and to the remark to the theorem (see Section 5.1.2) we see that some neighbourhood of the point is 1-1 projected onto each of the coordinate planes of the chosen coordinate system $Oxyz$.

Assume that the statement of the lemma is false. Then for every $\delta_n = 1/n$, $n = 1, 2, \ldots$ we can find a part S_n of S whose size is less than δ_n and which is not 1-1 projected onto the three coordinate planes of any Cartesian system. Choose in every part S_n a point M_n, and then choose a subsequence of the sequence $\{M_n\}$ converging to some point M of S. Consider the neighbourhood of M that is 1-1 projected onto each of the coordinate planes of some Cartesian system $Oxyz$. The neighbourhood contains one of the parts S_n that is also 1-1 projected onto the three coordinate planes of the system $Oxyz$. But this contradicts the choice of parts S_n. Thus the assumption that the statement of the lemma is false leads to contradiction. This proves the lemma.

Now we proceed *to prove Theorem* 7.3. Divide S with piecewise smooth curves into a finite number of smooth parts S_i, each of the size less than δ indicated in the lemma we have just proved. Add to the curves dividing S the edges of the surface. Since S_i is 1-1 projected onto the three coordinate planes of some Cartesian system, by the invariance of the Stokes formula (see Section 7.2.3) and the conclusions of Section 7.2.2 the Stokes formula is true for a part S_i. Now sum the left- and right-hand sides of the Stokes formula for the parts S_i. Obviously the sum of the left-hand sides of the formulas is the double integral $\iint\limits_S \boldsymbol{n} \operatorname{curl} \boldsymbol{p}\, d\sigma$ and the right-hand side is the sum of the integrals $\oint\limits_{\Gamma_i} \boldsymbol{p}t\, dl$ along the boundaries Γ_i of S_i. It is clear

* Such a part of the surface may be in a sphere of radius δ.

that the integrals along the common parts of the boundaries of S_i cancel out for they are traced in opposite directions (consider Fig. 7.6 for explanation). Therefore the above sum of line integrals

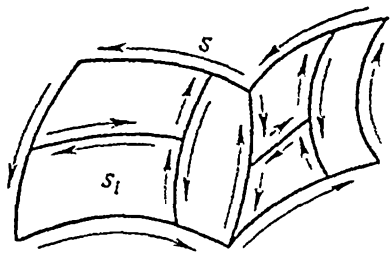


Fig. 7.6

is equal to the line integral along the boundary Γ of the surface S. Our reasoning implies the validity of the formula

$$\iint_S n \operatorname{curl} p \, d\sigma = \oint_\Gamma pt \, dl$$

which is just the Stokes formula. The proof of the theorem is complete

7.3. THE OSTROGRADSKY FORMULA

7.3.1. Statement of the main theorem. Let V be a finite, in general, multiply connected, domain with piecewise smooth boundary S* in the $Oxyz$ space. The domain V with adjoined boundary will be designated $\overline{V}$. The following *main* theorem is true.

Theorem 7.5. Let functions $P(x, y, z)$, $Q(x, y, z)$, *and* $R(x, y, z)$ *be continuous in* $\overline{V}$ *and have continuous partial derivatives of the first order in* V. *If there are improper integrals of each of the partial derivatives of* P, Q, *and* R *over* V, *we have the relation*

$$\iiint_V \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\, dy\, dz =$$

$$= \iint_S P\, dy\, dz + Q\, dz\, dx + R\, dx\, dy \qquad (7.27)$$

* The boundary S is said to be piecewise smooth if it is made up of a finite number of smooth surfaces adjoining one another along smooth curves, the surface edges. If S consists of a finite number of closed piecewise smooth surfaces S_i, then S_i are said to be *connected components of* S and the connected domain V is said to be *multiply connected*.

known as the **Ostrogradsky formula**. *The integral on the right of it is the sum of the integrals along the connected components of S on which the side exterior to V is chosen.*

We confine ourselves to proving the Ostrogradsky formula for only a special class of domains.

Note that Theorem 7.5 can be proved by extending the method used in Section 7.1 to prove the Green formula.

7.3.2. The proof of the Ostrogradsky formula for a special class of domains. A singly connected finite domain V is said to be a *type-K domain* if every straight line parallel to any coordinate axis intersects the boundary S of V in at most two points.

For the type-K domain special systems of exhaustive domains $\{\overline{V}_n\}$ are used. We now describe how such systems are constructed.

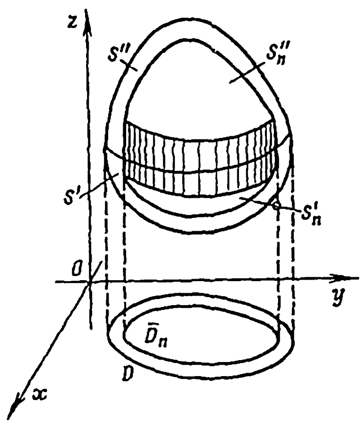


Fig. 7.7

Let a domain D in the Oxy plane be the projection of a domain V onto that plane. Draw through the boundary points of D straight lines parallel to the Oz axis. Each of the straight lines intersects the boundary S of V in one point only. The set of these points divides S into two parts, S' and S'', (Fig. 7.7) that are the graphs of the functions $z_1(x, y)$ and $z_2(x, y)$ continuous in $\overline{D}$ and piecewise differentiable in D. Note that $z_1(x, y) \leqslant z_2(x, y)$ (equality holding only at the points of the boundary of D).

Consider an arbitrary sequence of domains $\{\overline{D}_n\}$ monotonically exhausting D. Let S_n' and S_n'' be the graphs of the functions $z_1(x, y) + \varepsilon_n$ and $z_2(x, y) - \varepsilon_n$ given on $\overline{D}_n$ (ε_n is chosen to be so small that the surfaces S_n' and S_n'' do not intersect).

The boundary of $\overline{V}_n$ is the surface made up of the surfaces S_n' and S_n'' and part of a cylinder with generators parallel to the Oz axis, the boundary of a domain $\overline{D}_n$ serving as the directrix of the cylinder. The domain $\overline{V}_{n+1}$ is constructed similarly, but instead of $\overline{D}_n$ we take the domain $\overline{D}_{n+1}$ and choose ε_{n+1} to be less than ε_n. It is obvious that as $\varepsilon_n \to 0$ the system $\{\overline{V}_n\}$ monotonically exhausts the domain V.

We prove the following *statement*.

Theorem 7.6. In a type-K domain V, let functions $P(x, y, z)$, $Q(x, y, z)$, and $R(x, y, z)$ satisfy the hypotheses of Theorem 7.5. Then for that domain and for the functions P, Q, and R the Ostrogradsky formula is valid.

Proof. Clearly it suffices to show the validity of the equations

$$\iiint\limits_{V} \frac{\partial P}{\partial x}\,dx\,dy\,dz = \iint\limits_{S} P\,dy\,dz,$$

$$\iiint\limits_{V} \frac{\partial Q}{\partial y}\,dx\,dy\,dz = \iint\limits_{S} Q\,dz\,dx, \tag{7.28}$$

$$\iiint\limits_{V} \frac{\partial R}{\partial z}\,dx\,dy\,dz = \iint\limits_{S} R\,dx\,dy.$$

Since their proofs are of the same type we prove the third equation. Consider the triple integral

$$\iiint\limits_{\overline{V}_n} \frac{\partial R}{\partial z}\,dx\,dy\,dz. \tag{7.29}$$

For the domain $\overline{V}_n$ and for the integrand $\partial R/\partial z$ in the integral (7.29) all the conditions hold under which the repeated integration formula is valid. From this formula we have

$$\iiint\limits_{\overline{V}_n} \frac{\partial R}{\partial z}\,dx\,dy\,dz = \iint\limits_{\overline{D}_n} dx\,dy \int\limits_{z_1(x,\,y)+\varepsilon_n}^{z_2(x,\,y)-\varepsilon_n} \frac{\partial R}{\partial z}\,dz =$$

$$= \iint\limits_{\overline{D}_n} R(x,\ y,\ z_2(x,\ y)-\varepsilon_n)\,dx\,dy -$$

$$- \iint\limits_{\overline{D}_n} R(x,\ y,\ z_1(x,\ y)+\varepsilon_n))\,dx\,dy. \tag{7.30}$$

The left-hand side of relation (7.30) as $n \to 0$ has a limit equal to $\iiint\limits_{V} \frac{\partial R}{\partial z}\,dx\,dy\,dz$. By the uniform continuity of the function $R(x, y, z)$ in the closed domain $\overline{V}$ each of the terms on the right of (7.30) has as $n \to \infty$ a limit equal to $\iint\limits_{D} R(x, y, z_2(x, y))\,dx\,dy$ for the first term and to $\iint\limits_{D} R(x, y, z_1(x, y))\,dx\,dy$ for the second. The former of the integrals is (if the exterior side of S is chosen) the integral $\iint\limits_{S^*} R(x, y, z)\,dx\,dy$ and the latter (taking into account the "minus" sign preceding it) is the integral $\iint\limits_{S'} R(x, y, z)\,dx\,dy$. So the right-hand side of relations (7.30) has as $n \to \infty$ a limit

equal to $\iint\limits_S R(x, y, z)\, dx\, dy$. Consequently the third of the formulas (7.28) is proved.

The first and the second formulas (7.28) are proved similarly (it is necessary to consider the projections of V onto the Oyz and Oxz planes respectively and repeat the reasoning). The proof of the theorem is complete.

7.3.3. The invariant form of the Ostrogradsky formula. Let functions P, Q, and R satisfy the hypotheses of Theorem 7.5 in a finite connected domain V with piecewise smooth boundary S. We define in V a vector field $\boldsymbol{p}$ whose coordinates in a given Cartesian system $Oxyz$ are equal to P, Q, R. Under the conditions imposed on these functions the field $\boldsymbol{p}$ is clearly continuous in $\overline{V}$ and differentiable in V.

We find the divergence of $\boldsymbol{p}$. Using the expression for the divergence of $\boldsymbol{p}$ in an orthonormal basis $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ we get

$$\operatorname{div} \boldsymbol{p} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

Remark. We transform to a new Cartesian system in space. Let $\boldsymbol{i}'$, $\boldsymbol{j}'$, $\boldsymbol{k}'$ be an orthonormal basis referred to that system and let P', Q', R' be the coordinates of $\boldsymbol{p}$ in that basis. Obviously the functions P', Q', R' are continuous in $\overline{V}$ and differentiable in V (they are linear combinations of the functions P, Q, R).

Since in the new coordinate system

$$\operatorname{div} \boldsymbol{p} = \frac{\partial P'}{\partial x'} + \frac{\partial Q'}{\partial y'} + \frac{\partial R'}{\partial z'}.$$

by divergence invariance

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = \frac{\partial P'}{\partial x'} + \frac{\partial Q'}{\partial y'} + \frac{\partial R'}{\partial z'}.$$

Thus, if P, Q, R are considered as the coordinates of the vector field $\boldsymbol{p}$, the expression $\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$ remains unaltered in both form and value under a transformation to a new Cartesian system, i.e. is an invariant.

We may therefore make the following important conclusion: *the integral on the left of the Ostrogradsky formula* (7.27) *has an invariant nature, its form and value remaining unaltered under a transformation to a new Cartesian system.* Indeed, under such a transformation of coordinates the absolute value of the Jacobian is equal to unity. But according to the remark the integrand remains unaffected in both form and value under a transformation of coordinates.

We now look at the integral

$$\iint_S P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy \tag{7.31}$$

on the right of the Ostrogradsky formula (7.27). We shall show that *the integral also has an invariant nature, its form and the value of the integrand remaining unaffected under a transformation to a new Cartesian system.*

Using Remark 2 of Section 5.3.2 on the notation of the surface integral of the second kind and symbols X, Y, Z for the angles the normal n to the surface forms with the coordinate axes we may rewrite the integral (7.31) as

$$\iint_S (P\cos X + Q\cos Y + R\cos Z)\,d\sigma. \tag{7.32}$$

The integrand in the integral (7.32) is a scalar product np and the integral (7.32) (or equivalently the integral (7.31)) may therefore be written in the following *invariant form:*

$$\iint_S np\,d\sigma.$$

Note that this last integral is generally called the *flux of the vector field p* through the surface S.

Turning to the invariant notation of the integral (7.31) we see that in the new Cartesian system the integral has the form

$$\iint_S P'\,dy'\,dz' + Q'\,dz'\,dx' + R'\,dx'\,dy'.$$

Our arguments here allow the Ostrogradsky formula (7.31) to be written in the following invariant form:

$$\iiint_V \operatorname{div} p\,dv = \iint_S np\,d\sigma, \tag{7.33}$$

where dv denotes the volume element of V.

From Theorem 7.6 and conclusions of this subsection we may draw an important corollary.

Corollary. Let functions $P(x, y, z)$, $Q(x, y, z)$, and $R(x, y, z)$ satisfy the hypotheses of Theorem 7.5 in a finite domain V with piecewise smooth boundary S. If V may be divided into a finite number of domains V_k with piecewise smooth boundaries S_k, each of V_k a type-K domain referred to some Cartesian system, then for V and P, Q, and R the Ostrogradsky formula is true.

The validity of the corollary is obvious from the following reasoning. It is clear that the Ostrogradsky formula is true for

each of the domains V_k. This follows from the invariant nature of the formula and from Theorem 7.6 (in some coordinate system V_k is a type-K domain). Further, it is obvious that the sum of the integrals $\iiint\limits_{V_k} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\,dy\,dz$ from the left-hand sides of the Ostrogradsky formulas for the domains V_k is the integral $\iiint\limits_{V} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\,dy\,dz$. But the sum of the surface integrals $\iint\limits_{S_k} P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy$ on the right-hand sides of the Ostrogradsky formulas along the boundaries S_k of V_k gives $\iint\limits_{S} P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy$, for the integrals along the common parts of the boundaries of V_k cancel out, the parts in adjacent domains V_k being opposite in orientation.

7.4. SOME APPLICATIONS OF THE FORMULAS OF GREEN, STOKES AND OSTROGRADSKY

7.4.1. Expressing the area of a plane domain as a line integral. Let D be a finite plane connected domain with piecewise smooth boundary L. The following *statement* is true.

The area σ of D may be calculated from the formula

$$\sigma = \frac{1}{2} \oint\limits_{L} x\,dy - y\,dx \tag{7.34}$$

where the line integral is a sum of integrals along the connected components of L, a direction of circulation being indicated on each of the components such that D remains on the left.

To prove the statement, consider in D the functions

$$P(x, y) = -y, \quad Q(x, y) = x.$$

These clearly satisfy in D all conditions under which the Green formula (7.1) is true. From this formula we have

$$\iint\limits_{D} \left(\frac{\partial (x)}{\partial x} - \frac{\partial (-y)}{\partial y}\right) dx\,dy = \oint\limits_{L} (-y)\,dx + (x)\,dy.$$

The double integral in the last formula equals 2σ and the line integral equals $\oint\limits_{L} x\,dy - y\,dx$. Thus formula (7.34) is proved.

7.4.2. Expressing volume as a surface integral. Let V be a finite connected domain in space with piecewise smooth boundary S. The following *statement* is true.

The volume v of V can be calculated from the formula

$$v=\frac{1}{3}\iint_S x\,dy\,dz+y\,dz\,dx+z\,dx\,dy \tag{7.35}$$

in which the surface integral is a sum of integrals along the connected components of S, the side exterior to V being chosen on each of the components.

To prove the statement, consider in V the functions

$$P(x, y, z) = x, \quad Q(x, y, z) = y, \quad R(x, y, z) = z.$$

These clearly satisfy the conditions under which the Ostrogradsky formula is true. By this formula

$$\iiint_V \left(\frac{\partial (x)}{\partial x}+\frac{\partial (y)}{\partial y}+\frac{\partial (z)}{\partial z}\right) dx\,dy\,dz=$$

$$=\iint_S x\,dy\,dz+y\,dz\,dx+z\,dx\,dy.$$

The triple integral in the last formula equals $3v$. Therefore the last formula implies relation (7.35). This proves the statement.

7.4.3. Conditions under which the differential form $P(x, y)\,dx + Q(x, y)\,dy$ is a total differential. Here we show a number of conditions under which the differentiable form $P(x, y)\,dx + Q(x, y)\,dy$ given in the connected domain D is a total differential of some function $u(x, y)$.

We prove the following theorem.

Theorem 7.7. Let functions $P(x, y)$ and $Q(x, y)$ be continuous in a domain D. Then the following three conditions are equivalent.

1. *For any closed (possibly self-intersecting) piecewise smooth curve L in D*

$$\oint_L P\,dx+Q\,dy=0.$$

2. *For any two points A and B of D the value of the integral*

$$\int_{\breve{AB}} P\,dx+Q\,dy$$

is independent of the piecewise smooth curve $\breve{AB}$ joining the points A and B and lying in D.

3. *The differential form $P(x, y)\,dx + Q(x, y)\,dy$ is a total differential. In other words, a function $u(M) = u(x, y)$ is given in D such that*

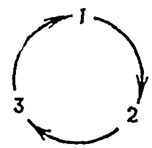
$$du = P\,dx + Q\,dy. \tag{7.36}$$

In this case, for any points A and B in D and for an arbitrary piecewise smooth curve AB joining those points and lying in D

$$\int_{\overset{\smile}{AB}} P\,dx + Q\,dy = u(B) - u(A). \tag{7.37}$$

Thus, fulfilling each of conditions 1, 2, 3 is necessary and sufficient for either of the other two to be fulfilled.

Proof. We prove the theorem using the diagram, i.e. prove that



the first condition implies the second, the second implies the third, and the third implies the first. It is obvious that this will prove the equivalence of conditions 1, 2, 3.

First step: $1 \to 2$. Let A and B be arbitrary fixed points in D, and let $\overset{\smile}{ACB}$ and $\overset{\smile}{AC'B}$ be any two piecewise smooth curves joining those points and lying in D (Fig. 7.8). The sum of the curves is a piecewise smooth (possibly self-intersecting) closed curve $L = \overset{\smile}{ACB} + \overset{\smile}{BC'A}$ in D. Since condition 1 is assumed fulfilled, we have

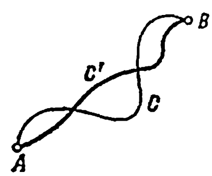


Fig. 7.8

$$\oint_L P\,dx + Q\,dy = 0.$$

From this, considering that $L = \overset{\smile}{ACB} + \overset{\smile}{BC'A}$ and that under a change of the direction of circulation the line integral changes sign, we get the relation

$$\int_{\overset{\smile}{ACB}} P\,dx + Q\,dy = \int_{\overset{\smile}{AC'B}} P\,dx + Q\,dy.$$

Consequently condition 2 holds.

Second step: $2 \to 3$. Let M_0 be a fixed point, let $M(x, y)$ be an arbitrary point of D, and let $\overset{\smile}{M_0M}$ be any piecewise smooth curve in D joining the points M_0 and M.

By condition 2 the expression

$$u(M) = \int_{\overset{\smile}{M_0M}} P\,dx + Q\,dy \tag{7.38}$$

is independent of $\overset{\frown}{M_0 M}$ and is therefore a function given in D. We prove that at each point M of D there are partial derivatives $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$, with

$$\frac{\partial u}{\partial x}=P(x, y), \quad \frac{\partial u}{\partial y}=Q(x, y). \tag{7.39}$$

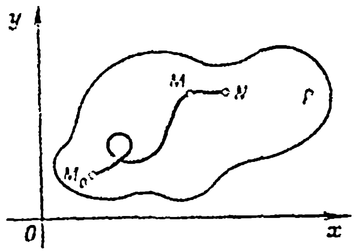


Fig. 7.9

Since $P(x, y)$ and $Q(x, y)$ are continuous in D, the last relations imply the differentiability of u and equation (7.36). This will prove the second step, $2 \to 3$.

The existence of partial derivatives of $u(x, y)$ is proved simultaneously with equations (7.39). We prove, for instance, the existence of $\partial u/\partial x$ and the first of the equations (7.39). Choose a point $M(x, y)$. Assign to the independent variable x an increment Δx small enough for the segment $\overline{MN}$ joining $M(x, y)$ and $N(x+\Delta x, y)$ to lie in D* (Fig. 7.9). We have

$$\Delta u = u(x+\Delta x, y)-u(x, y)=$$
$$=\int\limits_{\overset{\frown}{M_0 MN}} P\,dx+Q\,dy-\int\limits_{\overset{\frown}{M_0 M}} P\,dx+Q\,dy=\int\limits_{\overline{MN}} P\,dx+Q\,dy.$$

On $\overline{MN}$ the value of y is constant and therefore $\int\limits_{\overline{MN}} Q\,dy=0$. Consequently

$$\Delta u=\int\limits_{\overline{MN}} P\,dx=\int\limits_{x}^{x+\Delta x} P(t, y)\,dt.$$

Applying to the last integral the mean value theorem we get

$$\Delta u = P(x+\theta\Delta x, y)\,\Delta x, \text{ where } 0<\theta<1,$$

whence

$$\frac{\Delta u}{\Delta x}=P(x+\theta\Delta x, y), \quad 0<\theta<1.$$

By the continuity of $P(x, y)$ the right-hand side of the last equation has a limit as $\Delta x \to 0$ equal to the value of this function at $M(x, y)$. Consequently, the left-hand side too has the same limit equal by definition to $\frac{\partial u}{\partial x}$. The existence of a partial deriv-

* Since D is a domain, i.e. a set consisting of interior points only, such a choice of Δx is possible.

ative and the validity of the first of the equations (7.39) are thus proved. The existence of a partial derivative $\frac{\partial u}{\partial y}$ and the validity of the second of the equations (7.39) are proved similarly.

We now prove relation (7.37). Let A and B be any points in D, and let $\breve{AB}$ be an arbitrary piecewise smooth curve joining these points and lying in D. The curve is defined by the parametric equations $x = x(t)$, $y = y(t)$, $a \leqslant t \leqslant b$. Using the rule for evaluating line integrals we get

$$\int_{\breve{AB}} P\,dx + Q\,dy = \int_a^b \{P(x(t),\ y(t))\,x'(t) + Q(x(t),\ y(t))\,y'(t)\}\,dt =$$

$$= \int_a^b u_t'\,dt = u(x(b),\ y(b)) - u(x(a),\ y(a)) = u(B) - u(A).$$

Formula (7.37) is thus proved.

Third step: 3 $\to$ 1. This statement follows from formula (7.37). Indeed, for a closed curve L the end points coincide and therefore by formula (7.37)

$$\oint_L P\,dx + Q\,dy = u(A) - u(A) = 0.$$

The proof of the theorem is complete.

Remark. We noted that conditions 1, 2, 3 of Theorem 7.7 are equivalent and therefore, in particular, condition 3 is a necessary and sufficient condition under which the line integral $\int_L P\,dx + Q\,dy$ is independent of the choice of curve L joining any given points A and B in a domain D.

For singly connected domains* we shall show a necessary and sufficient condition, convenient for applications, that the differential form $P\,dx + Q\,dy$ should be a total differential of some function. Naturally this condition is necessary and sufficient for $\int_L P\,dx + Q\,dy$ to be independent of the choice of curve L joining any given points A and B in D.

Theorem 7.8. Let functions $P(x, y)$ and $Q(x, y)$ and their derivatives be continuous in a singly connected *domain D. Then each of the three conditions* 1, 2, 3 *of Theorem 7.7 is equivalent to*

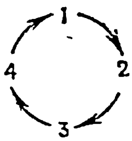
---

* Recall that a domain D is said to be singly connected if any piecewise smooth, not self-intersecting, closed curve in D bounds a domain all of whose points are in D.

condition

4. $\frac{\partial P}{\partial y}=\frac{\partial Q}{\partial x}$ *in* D.

Proof. We use the diagram



We have already proved the statements $1\to 2\to 3$. Now we prove that $3\to 4$ and $4\to 1$.

First step: $3\to 4$. Let there exist a function $u(x, y)$ in D such that $du=P\,dx+Q\,dy$. Then $\frac{\partial u}{\partial x}=P$, $\frac{\partial u}{\partial y}=Q$

$$\frac{\partial P}{\partial y}=\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right)=\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right)=\frac{\partial Q}{\partial x}.$$

Thus condition 4 holds. Note that proving the step $3\to 4$ does not require D to be singly connected.

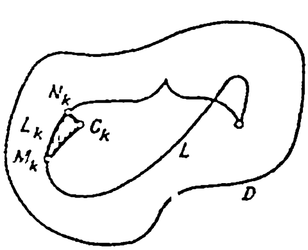


Fig. 7.10

Second step: $4\to 1$. Let condition 4 hold. Then at each point of D

$$\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}=0. \tag{7.40}$$

If L is a closed piecewise smooth curve without self-intersection lying in D and bounding a domain D^* (D is singly connected and therefore each point of D^* is in D), then applying the Green formula to D^* and using (7.40) we get

$$\oint_L P\,dx+Q\,dy=\iint_{D^*}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dx\,dy=0.$$

When L has a finite number of points of self-intersection or is a broken line with a finite number of components, $\oint_{\widetilde{L}} P\,dx+Q\,dy=0$ for each loop $\widetilde{L}$ of L and therefore $\oint_L P\,dx+Q\,dy=0$ for L.

Let L be an arbitrary closed piecewise smooth curve. Choose for L a number $\lambda>0$ in the way indicated in Lemma 1. Divide L into components L_k of length less than λ (the corner points of L are also among the division points, see Fig. 7.10). According to Lemma 1 the tangents at the end points M_k and N_k of each of the components

L_k make an angle less than $\pi/8$. Then obviously for a sufficiently small λ the curvilinear triangle $M_kN_kC_k$ (the shaded triangle in Fig. 7.10), in which M_kC_k makes with the tangent at M_k an angle less than $\pi/8$ and N_kC_k is the normal to L at N_k, is entirely in D and is a closed piecewise smooth curve without self-intersections. Therefore

$$\oint\limits_{M_kN_kC_k} P\,dx+Q\,dy=0.$$

It follows that the line integral along the arc $\overset{\frown}{M_kN_k}$ is equal to the line integral along the broken line $M_kC_kN_k$:

$$\int\limits_{\overset{\frown}{M_kN_k}} P\,dx+Q\,dy=\int\limits_{M_kC_kN_k} P\,dx+Q\,dy.$$

Similarly for any component L_k we obtain in D a closed broken line $\hat{L}$ for which

$$\oint\limits_{\hat{L}} P\,dx+Q\,dy=\oint\limits_{L} P\,dx+Q\,dy. \tag{7.41}$$

We noted above that for a closed broken line $\hat{L}$ in D the integral $\int\limits_{\hat{L}} P\,dx+Q\,dy=0$. From this and from (7.41) we get

$$\oint\limits_{L} P\,dx+Q\,dy=0.$$

The proof of the theorem is complete.

7.4.4. Potential and solenoidal vector fields. Earlier (see Sections 7.1.3, 7.2.3, and 7.3.3) we introduced the concepts of circulation and flux of a vector field. We recall these notions.

Let a continuous vector field $\mathbf{p}(M)=\mathbf{p}(x, y, z)$ be given in some domain D.

Definition 1. The circulation of a vector field $\mathbf{p}$ *about a closed piecewise smooth curve* L *lying in* D *is the integral*

$$\oint\limits_{L} \mathbf{pt}\,dl,$$

where $\mathbf{t}$ *is a unit tangent vector to* L *and* dl *is the differential of the arc length of* L.

Definition 2. The flux of a vector field p through an oriented piecewise smooth surface S in D is the integral

$$\iint_S pn\,d\sigma,$$

where n is a unit normal vector to S indicating the orientation of S and $d\sigma$ is the area element of S.

We introduce the concepts of *potential* and *solenoidal* vector fields.

Definition 3. A vector field p is said to be potential *in D if the circulation of the field about any closed piecewise smooth curve in D is zero.*

Definition 4. A vector field p is said to be solenoidal *in D if the flux of p through any piecewise smooth, not self-intersecting closed surface in D that is the boundary of some bounded subdomain of D is zero.*

For continuous differentiable vector fields and a special class of domains we shall prove a theorem containing necessary and sufficient conditions for a field to be potential.

As a preliminary we shall introduce the concept of *three-dimensional singly-connected-surface domain* or *SCS domain.*

A three-dimensional domain D is said to be a *singly-connected-surface domain* (or *SCS domain*) if for any piecewise smooth closed curve L in D we can find an orientable piecewise smooth surface S in D such that its boundary is L. Note that for that surface S the Stokes formula is true.

The following theorem holds.

Theorem 7.9. Let a continuously differentiable vector field $p = \{P, Q, R\}$ be given in a SCS domain D. Then the following three conditions are equivalent:

1. *The vector field $p = p(M)$ is potential.*
2. *In D there is a potential function $u(M)$, i.e. a function such that $p = \operatorname{grad} u$ or equivalently*

$$du = P\,dx + Q\,dy + R\,dz.$$

In this case for any points A and B in D and for an arbitrary piecewise smooth curve AB joining these points and lying in D

$$\int_{\breve{AB}} pt\,dl = u(b) - u(A),$$

where t is a unit tangent vector to AB and dl is the differential of an arc.

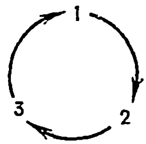
3. *The vector field $p = p(M)$ is irrotational, i.e.* $\operatorname{curl} p = 0$ *in D.*

Condition 3 is obviously equivalent to the relations

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}, \quad \frac{\partial Q}{\partial z} = \frac{\partial R}{\partial y}, \quad \frac{\partial R}{\partial x} = \frac{\partial P}{\partial z}.$$

Either of conditions 2 and 3 is thus a necessary and sufficient condition for the differentiable vector field p to be potential.

Proof. We use the diagram



The statements $1 \to 2$ and $2 \to 3$ are true without assuming D to be a SCS domain and their proofs are quite similar to those of the corresponding statements of Theorems 7.7 and 7.8.

We prove the statement $3 \to 1$.

Let L be a piecewise smooth closed curve in D. Under the hypothesis D is a SCS domain. There is therefore a piecewise smooth surface S in D such that its boundary is L. By the Stokes formula (7.26)

$$\oint_L \boldsymbol{p}t\,dl = \iint_S \boldsymbol{n} \operatorname{curl} \boldsymbol{p}\,d\sigma .$$

From this and from the condition $\operatorname{curl} \boldsymbol{p} = 0$ we get

$$\oint_L \boldsymbol{p}t\,dl = 0,$$

i.e. the field $\boldsymbol{p}$ is potential. The proof of the theorem is complete.

In conclusion we prove the theorem on necessary and sufficient conditions for a vector field in the so-called *singly-connected-volume domains* (or *SCV domains*) to be solenoidal. A spatial domain D is said to be a singly-connected-volume domain or SCV domain if any piecewise smooth, not self-intersecting closed orientable surface in D is the boundary of a domain that is also in D.

Theorem 7.10. For a continuously differentiable vector field $\boldsymbol{p}$ *to be solenoidal in a SCV domain* D *it is necessary and sufficient that at every point of* D

$$\operatorname{div} \boldsymbol{p} = 0.$$

Proof. (1) *Necessity.* Let M be an arbitrary point of D. Consider any sphere S with centre at M lying entirely in D. Applying to the ball D_S with boundary S the Ostrogradsky formula (7.33) we get

$$\iiint_{D_S} \operatorname{div} \boldsymbol{p}\,dv = \iint_S \boldsymbol{np}\,d\sigma. \tag{7.42}$$

Since the field $\boldsymbol{p}$ is solenoidal, $\iint_S \boldsymbol{np}\,d\sigma = 0$ and therefore by (7.42) $\iiint_{D_S} \operatorname{div} \boldsymbol{p}\,dv = 0$. Applying to the last integral the mean value theorem we see that at some point of D_S $\operatorname{div} \boldsymbol{p} = 0$. By virtue of

the arbitrariness of the ball and the continuity of the field p this implies that div p vanishes at M. The necessity of the hypotheses of the theorem is thus proved.

(2) *Sufficiency*. Let S be any closed, piecewise smooth, not self-intersecting, orientable surface in D. Since D is a SCV domain, S is the boundary of the domain D_S that is also in D. Applying to D_S and to p the Ostrogradsky formula (7.33) we obtain the relation

$$\iint_S np\, d\sigma = 0.$$

Since S is an arbitrary piecewise smooth, not self-intersecting closed orientable surface in D, the last equation implies by definition that the field p is solenoidal in D. The proof of the theorem is complete.

SUPPLEMENT

DIFFERENTIAL FORMS IN EUCLIDEAN SPACE

7S.1. Skew-Symmetric Multilinear Forms

7S.1.1. Linear forms. Let V be an arbitrary n-dimensional vector space whose elements are designated $\xi, \eta, \ldots$. We shall study functions associating with each element $\xi \in V$ some real number.

Definition 1. A function $a(\xi)$ *is said to be a* **linear form** *if for any* $\xi \in V$, $\eta \in V$ *and any real number* λ

(1) $a(\xi + \eta) = a(\xi) + a(\eta)$,

(2) $a(\lambda\xi) = \lambda a(\xi)$.

Definition 2. A sum of two linear forms a *and* b *is a linear form associating with each vector* $\xi \in V$ *a number*

$$c(\xi) = a(\xi) + b(\xi).$$

A product of a linear form a *by a real number* λ *is a linear form* b *associating with each vector* $\xi \in V$ *a number*

$$b(\xi) = \lambda a(\xi).$$

The set of all linear forms thus forms a vector space which we designate $L(V)$*. We find the representation of the linear form a in some basis $\{e_i\}_{i=1}^n$. Let

$$\xi = \sum_{i=1}^{n} \xi^i e_i,$$

where the numbers ξ^i are determined uniquely. Denoting $a_i = a(e_i)$, the desired representation has the form

$$a(\xi) = \sum_{i=1}^{n} \xi^i a_i.$$

* The space $L(V)$ is also designated V^* and called *conjugate* (or *dual*) to V.

7S.1.3. Multilinear forms. Let p be a natural number. We use $V^p = V \times V \times \times \ldots \times V$ to denote the set of all ordered collections $(\xi_1, \xi_2, \ldots, \xi_p)$ of p vectors, each lying in V, and consider the functions associating with each of such collections some real number.

Definition. A function $a(\xi_1, \xi_2, \ldots, \xi_p)$ is said to be **a multilinear form of degree** p *(or* p**-form**) *if it is a linear form in each of the independent variables with the values of the rest fixed.*

Introducing linear operations into the set of all p-forms we obtain a linear space which we designate $L_p(V)$.

We find the representation of an arbitrary multilinear form $a(\xi_1, \xi_2, \ldots, \xi_p)$ in some basis $\{e_i\}_{i=1}^n$ of V. We denote

$$a_{i_1 i_2 \ldots i_p} = a(e_{i_1}, e_{i_2}, \ldots, e_{i_p}).$$

Hence if $\xi_k = \sum_{i=1}^{n} \xi_k^i e_i$, $k = 1, 2, \ldots, p$, then

$$a(\xi_1, \xi_2, \ldots, \xi_p) = \sum_{i_1=1}^{p} \ldots \sum_{i_p=1}^{p} a_{i_1 i_2 \ldots i_p} \xi_1^{i_1} \xi_2^{i_2} \ldots \xi_p^{i_p}.$$

If $e^k(\xi)$ is a basis in $L(V)$ conjugate to $\{e_i\}$, then obviously the p-forms

$$e^{i_1 i_2 \ldots i_p}(\xi_1, \xi_2, \ldots, \xi_p) = e^{i_1}(\xi_1) e^{i_2}(\xi_2) \ldots e^{i_p}(\xi_p)$$

form a basis in $L_p(V)$ and thus $L_p(V)$ is of dimension n^p.

7S.1.4. Skew-symmetric multilinear forms.

Definition. A multilinear form $a(\xi_1, \xi_2, \ldots, \xi_p)$ is said to be **skew-symmetric** *if under a permutation of any two independent variables it changes sign*. In other words*

$$a(\xi_1, \xi_2, \ldots, \xi_i, \ldots, \xi_j, \ldots, \xi_p) = -a(\xi_1, \xi_2, \ldots, \xi_j, \ldots, \xi_i, \ldots, \xi_p).$$

Obviously, a set of all multilinear skew-symmetric forms of degree p form a subspace of $L_p(V)$ which we designate $A_p(V)$**. The elements of $A_p(V)$ are designated $\omega = \omega(\xi_1, \xi_2, \ldots, \xi_p)$.

Notice that if $\{e_i\}$ is an arbitrary basis in V and

$$\omega = \sum_{i_1=1}^{n} \ldots \sum_{i_p=1}^{n} \omega_{i_1 \ldots i_p} \xi_1^{i_1} \ldots \xi_p^{i_p},$$

then the numbers $\omega_{i_1 \ldots i_p}$ change sign under a permutation of two indices. This follows from the fact that

$$\omega_{i_1 \ldots i_p} = \omega(e_{i_1}, \ldots, e_{i_p}).$$

It is natural to consider that $A_1(V) = L_1(V)$ and that $A_0(V)$ consists of all constants, i.e. coincides with the number line.

7S.1.5. External product of skew-symmetric forms. Consider two skew-symmetric forms $\omega^p \in A_p(V)$ and $\omega^q \in A_q(V)$. We introduce here a basic operation in the theory of skew-symmetric forms, the operation of *external multiplication*.

---

* Skew-symmetric multilinear forms are also called *antisymmetric*, *skew*, or *exterior*.

** This space is also designated $\wedge^p V^*$ and called the pth exterior degree of V^*.

Let

$\omega^p = \omega^p(\eta_1, \eta_2, \ldots, \eta_p), \quad \eta_i \in V,$

$\omega^q = \omega^q(\zeta_1, \zeta_2, \ldots, \zeta_q), \quad \zeta_j \in V.$

Consider the following multilinear form $a = L_{p+q}(V)$:

$$a(\xi_1, \xi_2, \ldots, \xi_{p+q}) = \omega^p(\xi_1, \ldots, \xi_p)\cdot\omega^q(\xi_{p+1}, \ldots, \xi_{p+q}). \tag{7.43}$$

This in general is not skew-symmetric. That is, under a permutation of ξ_i and ξ_j, where $1 \leqslant i \leqslant p$ and $p+1 \leqslant j \leqslant p+q$, the form (7.43) may not change sign. It is this that necessitates the introduction of external product.

To introduce external product we shall need some facts from permutation theory.

Recall that a *permutation* of numbers $\{1, 2, \ldots, m\}$ is the function $\sigma = \sigma(k)$ defined on these numbers and mapping them in a one-to-one manner onto itself. The set of all such permutations is designed $\sum_m$. Obviously there are in all $m!$ different permutations in $\sum_m$. For two permutations $\sigma \in \sum_m$ and $\tau \in \sum_m$ a superposition $\sigma\tau \in \sum_m$ is uniquely defined. The permutation σ^{-1} is the inverse of σ if $\sigma^{-1}\sigma = \sigma\sigma^{-1} = \varepsilon$, where ε is an identity permutation (i.e. $\varepsilon(k) = k$, $k = 1, 2, \ldots, m$).

A permutation σ is said to be a *transposition* if it interchanges two numbers, leaving the others fixed. In other words, there is a pair of numbers i and j ($1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant m$, $i \neq j$) such that $\sigma(i) = j$, $\sigma(j) = i$, and $\sigma(k) = k$ for $k \neq i$ and $k \neq j$. Clearly, if σ is a transposition, then $\sigma^{-1} = \sigma$ and $\sigma\cdot\sigma = \varepsilon$.

It is known that any permutation σ can be represented as a superposition of transpositions interchanging adjacent numbers, the parity of the number of transpositions in such a decomposition being independent of the choice of decomposition and is called the parity of the permutation σ.

We introduce the following notation:

$$\operatorname{sgn}\sigma = \begin{cases} 1, & \text{if } \sigma \text{ is even,} \\ -1, & \text{if } \sigma \text{ is odd.} \end{cases}$$

Notice that a form $a \in L_p(V)$ belongs to $A_p(V)$ if for any permutation $\sigma \in \sum_p$

$$a(\xi_{\sigma(1)}, \xi_{\sigma(2)}, \ldots, \xi_{\sigma(p)}) = \operatorname{sgn}\sigma\cdot a(\xi_1, \xi_2, \ldots \xi_p).$$

Consider again the multilinear form (7.43). For any permutation $\sigma \in \sum_{p+q}$ we set

$$\sigma a(\xi_1, \ldots, \xi_{p+q}) = a(\xi_{\sigma(1)}, \ldots, \xi_{\sigma(p+q)}). \tag{7.44}$$

It is easily seen that if $\tau \in \sum_{p+q}$ and $\sigma \in \sum_{p+q}$, then $(\tau\sigma)a = \tau(\sigma a)$.

We introduce the following definition.

Definition. *The external product of a form $\omega^p \in A_p(V)$ and a form $\omega^q \in A_q(V)$ is a form $\omega \in A_{p+q}(V)$ defined by the equation*

$$\omega(\xi_1, \ldots, \xi_{p+q}) = \sum_\sigma \operatorname{sgn}\sigma\cdot\sigma a, \tag{7.45}$$

where the sum is taken over all permutations $\sigma \in \sum_{p+q}$ satisfying the condition

$$\sigma(1) < \sigma(2) < \ldots < \sigma(p), \quad \sigma(p+1) < \ldots < \sigma(p+q) \tag{7.46}$$

and the value of σa is given by equations (7.43) and (7.44).

The external product of ω^p and ω^q is designated

$$\omega = \omega^p \wedge \omega^q$$

By way of illustration, consider the action of a permutation σ satisfying condition (7.46). Suppose two columns of vehicles are moving in parallel along some road, the first column consisting of p vehicles and the second of q vehicles. After some time the road becomes narrower and the two columns rearrange into one column without stopping. The vehicles of the first column occupy places somewhere among the vehicles of the second, but the sequence of vehicles within either column remains the same. As a result, we obtain a permutation satisfying condition (7.46). It is easy to see that, conversely, any such permutation can be realized in our model.

To see that our definition is correct it is necessary to prove that $\omega = \omega^p \wedge \omega^q \in A_{p+q}(V)$. Clearly it is only necessary to prove the antisymmetry of the form ω.

We show that under a permutation of two independent variables ξ_i and ξ_{i+1} the form ω *changes sign*. It will then easily follow that $\omega \in A_{p+q}(V)$. Let $\tau \in \sum_{p+q}$ be such a permutation. We shall see that

$$\tau\omega = -\omega = (\operatorname{sgn} \tau)\, \omega. \tag{7.47}$$

From (7.45) we get

$$\tau\omega = \sum_{\sigma} (\operatorname{sgn} \sigma)\cdot(\tau\sigma)\, a.$$

We split this sum into two:

$$\tau\omega = \sum_{\sigma}{}' (\operatorname{sgn} \sigma)(\tau\sigma)\, a + \sum_{\sigma}{}'' (\operatorname{sgn} \sigma)(\tau\sigma)\, a. \tag{7.48}$$

The *first sum* will contain those permutations σ for which either $\sigma^{-1}(i) \leqslant p$, $\sigma^{-1}(i+1) \leqslant p$ or $\sigma^{-1}(i) \geqslant p+1$, $\sigma^{-1}(i+1) \geqslant p+1$. For each of such permutations

$$(\tau\sigma)\, a = -\sigma a.$$

To make this statement clearer, we denote $k = \sigma^{-1}(i)$, $l = \sigma^{-1}(i+1)$, i.e. $i = \sigma(k)$, $i+1 = \sigma(l)$. The form σa is a product of ω^p and ω^q, the independent variables of ω^p being the vectors $\xi_{\sigma(1)}, \xi_{\sigma(2)}, \ldots, \xi_{\sigma(p)}$, and of ω^q the vectors $\xi_{\sigma(p+1)}, \ldots, \xi_{\sigma(p+q)}$. If $k \leqslant p$ and $l \leqslant p$, then $\xi_i = \xi_{\sigma(k)}$ and $\xi_{i+1} = \xi_{\sigma(l)}$ are the independent variables of the form ω^p which is skew-symmetric as stated. Consequently, interchanging ξ_i and ξ_{i+1} makes ω^p and hence σa change sign. The case where $k \geqslant p+1$ and $l \geqslant p+1$ is considered along the same lines.

So for the first sum we have

$$\sum_{\sigma}{}' (\operatorname{sgn} \sigma)(\tau\sigma)\, a = -\sum_{\sigma}{}' (\operatorname{sgn} \sigma)\, \sigma a. \tag{7.49}$$

The *second sum* will contain those permutations σ for which either $\sigma^{-1}(i) \leqslant p$, $\sigma^{-1}(i+1) \geqslant p+1$ or $\sigma^{-1}(i) \geqslant p+1$, $\sigma^{-1}(i+1) \leqslant p$. We show that the set of permutations $\{\sigma\}$ satisfying this condition (and of course condition (7.46) as well) coincides with the set of permutations of the form $\tau\sigma$, where $\sigma \in \{\sigma\}$. Let us turn to our two-column model. The statement will assume the following obvious form.

If under any rearrangement a vehicle k of the first column finds itself directly in front of a vehicle l of the second column, then we can easily find another rearrangement as a result of which those vehicles would interchange places, the sequence of the other vehicles remaining unaltered.

Thus, since $\operatorname{sgn} \tau\sigma = -\operatorname{sgn} \sigma$

$$\sum_{\sigma}{}'' (\operatorname{sgn} \sigma)(\tau\sigma)\, a = -\sum_{\sigma}{}'' (\operatorname{sgn} \tau\sigma)(\tau\sigma)\, a = -\sum_{\sigma}{}'' (\operatorname{sgn} \sigma)\, \sigma a. \tag{7.50}$$

Substituting (7.49) and (7.50) in (7.48) we obtain (7.47).

Example 1. Consider two linear forms $f(\xi) \in A_1(V)$ and $g(\xi) \in A_1(V)$. The external product is the bilinear form

$$f \wedge g = \sum_{\sigma} (\operatorname{sgn} \sigma) \cdot \sigma f(\xi_1) g(\xi_2) = f(\xi_1) g(\xi_2) - g(\xi_1) f(\xi_2).$$

Example 2. Let $f(\xi) \in A_1(V)$, $g(\xi_1, \xi_2, \ldots, \xi_q) \in A_q(V)$. The external product $\omega = f \wedge g$ is a $q+1$-form whose independent variables we denote by $\xi_0, \xi_1, \ldots, \xi_q$

$$\omega = \sum_{\sigma} (\operatorname{sgn} \sigma) \sigma f(\xi_0) g(\xi_1, \xi_2, \ldots, \xi_q) =$$

$$= \sum_{i=0}^{q} (-1)^i f(\xi_i) g(\xi_0, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_q).$$

7S.1.6. **Properties of external product of skew-symmetric forms.** (1) An obvious property of external product is *linearity*:

(a) if $\omega^p \in A_p(V)$, $\omega^q \in A_q(V)$, then for any real number λ

$$(\lambda\omega^p) \wedge \omega^q = \omega^p \wedge (\lambda\omega^q) = \lambda(\omega^p \wedge \omega^q);$$

(b) if $\omega_1^p \in A_p(V)$, $\omega_2^p \in A_p(V)$ and $\omega^q \in A_q(V)$, then

$$(\omega_1^p + \omega_2^p) \wedge \omega^q = \omega_1^p \wedge \omega^q + \omega_2^p \wedge \omega^q.$$

(2) *Anticommutativity*. If $\omega^p \in A_p(V)$ and $\omega^q \in A_q(V)$, then

$$\omega^p \wedge \omega^q = (-1)^{pq} \omega^q \wedge \omega^p.$$

Proof. Let

$$\omega^p \wedge \omega^q = \omega = \omega(\xi_1, \xi_2, \ldots, \xi_{p+q}).$$

It is easy to see that

$$\omega^q \wedge \omega^p = \omega(\xi_{p+1}, \xi_{p+2}, \ldots, \xi_{p+q}, \xi_1, \ldots, \xi_p).$$

We show that the permutation $(\xi_{p+1}, \ldots, \xi_{p+q}, \xi_1, \ldots, \xi_p)$ may be obtained from the vectors $(\xi_1, \ldots, \xi_{p+q})$ by means of pq successive transpositions. The vector ξ_{p+1} may be moved to the first place using p transpositions. Using the same number of transpositions we may then move to the second place the vector ξ_{p+2}, and so on. In all, we shall move p vectors, using p transpositions each time, i.e. the number of all transpositions is pq. In this case anticommutativity is implied by the antisymmetry of the external product.

(3) *Associativity*. If $\omega^p \in A_p(V)$, $\omega^q \in A_q(V)$, $\omega^r \in A_r(V)$, then

$$(\omega^p \wedge \omega^q) \wedge \omega^r = \omega^p \wedge (\omega^q \wedge \omega^r).$$

Proof. Let $\sigma \in \sum_{p+q+r}$. Consider

$$\omega = \sum_{\sigma} (\operatorname{sgn} \sigma) \sigma [\omega^p(\xi_1, \ldots, \xi_p) \omega^q(\xi_{p+1}, \ldots, \xi_{p+q}) \omega^r(\xi_{p+q+1}, \ldots, \xi_{p+q+r})]. \tag{7.51}$$

The sum (7.51) equals $(\omega^p \wedge \omega^q) \wedge \omega^r$ if we first sum over all permutations that leave unaltered the numbers $p+q+1$, $p+q+2, \ldots, p+q+r$ and satisfy condition (7.46) and then sum over all permutations that preserve the resulting order of the first $p+q$ independent variables and the order of the variables $\xi_{p+q+1}, \ldots, \xi_{p+q+r}$.

Similarly we can obtain $\omega^p \wedge (\omega^q \wedge \omega^r)$.

We show that in both cases we obtain a sum over all permutations satisfying the conditions

$$\left.\begin{aligned} \sigma(1) < \sigma(2) < \ldots < \sigma(p), \\ \sigma(p+1) < \sigma(p+2) < \ldots < \sigma(p+q), \\ \sigma(p+q+1) < \ldots < \sigma(p+q+r). \end{aligned}\right\} \tag{7.52}$$

To do this we turn again to our vehicle-column model. Suppose there are three columns of vehicles moving along the road, p vehicles in one, q in another, and r in the third. One of the ways to rearrange the three columns into one is to first merge the first and the second column and then join the resulting column to the third. Alternatively, the second and third columns may first be merged and the first joined to them. Obviously the permutation σ resulting from either of these rearrangements satisfies condition (7.52) and conversely either permutation satisfying condition (7.52) may be obtained using either the first or the second way of rearranging the vehicles. This just means that $(\omega^p \wedge \omega^q) \wedge \omega^r$ and $\omega^p \wedge (\omega^q \wedge \omega^r)$ coincide.

The associativity of external product enables one to consider any finite product

$$\omega_1 \wedge \omega_2 \wedge \ldots \wedge \omega_m, \text{ where } \omega_l \in A_{p_l}(V).$$

Exampe 1. Let $a_1(\xi)$, $a_2(\xi)$, $\ldots$, $a_m(\xi)$ be linear forms. Then

$$a_1 \wedge a_2 \wedge \ldots \wedge a_m = \sum_{\sigma} (\operatorname{sgn} \sigma)\, \sigma\, [a_1(\xi_1)\, a_2(\xi_2) \ldots a_m(\xi_m)], \tag{7.53}$$

where the summation is taken over all permutations $\sigma \in \sum_m$.

The equation is easy to verify by induction. Notice that if we introduce the matrix $\{a_i(\xi_j)\}$, then equation (7.53) may be rearranged in the form

$$(a_1 \wedge a_2 \wedge \ldots \wedge a_m)(\xi_1, \xi_2, \ldots, \xi_m) = \det\{a_i(\xi_j)\}. \tag{7.54}$$

7S.1.7. Basis in space of skew-symmetric forms. Choose some basis $\{e_i\}_{i=1}^n$ in a space V and denote by $\{e^i\}_{i=1}^n$ the conjugate basis in $L(V)$. Recall that $e^i(\xi)$ is a linear form which on the elements of the basis $\{e_j\}$ takes the value $e^i(e_j) = \delta_{ij}$.

In Section 7S.1.3 we showed that all possible products

$$e^{i_1}(\xi_1)\, e^{i_2}(\xi_2) \ldots e^{i_p}(\xi_p)$$

form a basis in $L_p(V)$. Since $A_p(V) \subset L_p(V)$, every skew-symmetric p-form may be decomposed uniquely as a linear combination of those products. However, they form no basis in $A_p(V)$, since they are not skew-symmetric p-forms, i.e. do not belong to $A_p(V)$. Nevertheless we can construct from them a basis in $A_p(V)$ using external multiplication.

Theorem 7.11. Let $\{e_i\}_{i=1}^n$ be a basis in a space V, and let $\{e^i\}_{i=1}^n$ be the conjugate basis in $L(V)$. Any skew-symmetric p-form $\omega \in A_p(V)$ may be represented, uniquely, as

$$\omega = \sum_{1 \le i_1 < \ldots < i_p \le n} \omega_{i_1 i_2 \ldots i_p} e^{i_1} \wedge e^{i_2} \wedge \ldots \wedge e^{i_p}. \tag{7.55}$$

Each term on the right side of (7.55) is a product of a constant $\omega_{i_1 i_2 \ldots i_p}$ by a skew-symmetric p-form $e^{i_1} \wedge e^{i_2} \wedge \ldots \wedge e^{i_p}$.

Proof. By virtue of the results of Sec. 7S.1.4 we may write

$$\omega = \sum_{i_1=1}^{n} \ldots \sum_{i_p=1}^{n} \omega_{i_1 i_2 \ldots i_p} e^{i_1} e^{i_2} \ldots e^{i_p}, \tag{7.56}$$

where the numbers $\omega_{i_1 i_2 \ldots i_p} = \omega(e^{i_1}, e^{i_2}, \ldots, e^{i_p})$ are determined uniquely.

Since the form $\omega(\xi_1, \xi_2, \ldots, \xi_p)$ is skew-symmetric, for any permutation $\sigma \in \sum_p$

$$\omega(\xi_{\sigma(1)}, \xi_{\sigma(2)}, \ldots, \xi_{\sigma(p)}) = (\operatorname{sgn} \sigma)\, \omega(\xi_1, \xi_2, \ldots, \xi_p).$$

Hence

$$\omega_{i_{\sigma(1)} i_{\sigma(2)} \ldots i_{\sigma(p)}} = (\operatorname{sgn} \sigma)\, \omega_{i_1 i_2 \ldots i_p}. \tag{7.57}$$

Group together the terms in (7.56) that differ in permutation of the indices $i_1, i_2, \ldots, i_p$ and use equation (7.57). We get

$$\omega = \sum_{i_1 < i_2 < \ldots < i_p} \sum_{\sigma} \omega_{i_{\sigma(1)} \ldots i_{\sigma(p)}} e^{i_{\sigma(1)}} \ldots e^{i_{\sigma(p)}} =$$

$$= \sum_{i_1 < i_2 < \ldots < i_p} \omega_{i_1 i_2 \ldots i_p} \Big[\sum_{\sigma} (\operatorname{sgn} \sigma)\, e^{i_{\sigma(1)}} \ldots e^{i_{\sigma(p)}} \Big]. \tag{7.58}$$

By virtue of the example in Sec. 7S.1.6 the sum in square brackets is $e^{i_1} \wedge e^{i_2} \wedge \ldots \wedge e^{i_p}$. The proof of the theorem is complete.

Corollary 1. The elements $e^{i_1} \wedge e^{i_2} \wedge \ldots \wedge e^{i_p}$ $(1 \leqslant i_1 < i_2 < \ldots i_p \leqslant n)$ *form a basis in* $A_p(V)$. *That basis is empty for* $p > n$ *and consists of one element if* $p = n$.

Corollary 2. The dimensionality of the space $A_p(V)$ *is* C_n^p.

In what follows we shall assume as a rule that the chosen basis $e_1, e_2, \ldots, e_n$ is fixed and designate the linear forms $e^i(\xi)$ as $e^i(\xi) = \xi^i$. Then any form $\omega \in A_p(V)$ becomes

$$\omega(\xi_1, \xi_2, \ldots, \xi_p) = \sum_{i_1 < \ldots < i_p} \omega_{i_1 \ldots i_p} \xi^{i_1} \wedge \ldots \wedge \xi^{i_p}. \tag{7.59}$$

Example 1.

$$\xi^1 \wedge \xi^2 = (e^1 \wedge e^2)(\xi_1, \xi_2) = \sum_{\sigma} (\operatorname{sgn} \sigma)\, \sigma\, [e^1(\xi_1)\, e^2(\xi_2)] =$$

$$= e^1(\xi_1)\, e^2(\xi_2) - e^1(\xi_2)\, e^2(\xi_1) = \xi_1^1 \xi_2^2 - \xi_2^1 \xi_1^2,$$

where ξ_i^j is the jth coefficient in the expansion of a vector ξ_i with respect to the basis $\{e_j\}$.

Example 2.

$$\xi^1 \wedge \xi^2 \wedge \ldots \wedge \xi^n = \det\{\xi_i^j\},$$

where $\xi_i = \sum_{j=1}^{n} \xi_i^j e_j$.

7S.2. Differential Forms

7S.2.1. Definitions. Consider an arbitrary open domain G of an n-dimensional Euclidean space E^n. The points of G are designated $x = (x^1, x^2, \ldots, x^n)$, $y = (y^1, y^2, \ldots, y^n)$, and so on.

Definition. A differential form of degree p defined in G is a function $\omega(x, \xi_1, \xi_2, \ldots, \xi_p)$ which is a skew-symmetric p-form in $A_p(E^n)$ for each fixed $x \in G$.

The set of all differential p-forms in G is denoted by $\Omega_p(G) = \Omega_p(G, E^n)$.

We shall assume that for fixed $\xi_1, \ldots, \xi_p \in E^n$ a p-form ω is a function infinitely differentiable in G. Using the results of Section 7S.1 we may write every p-form ω as

$$\omega = \sum_{i_1<\ldots<i_p} \omega_{i_1\ldots i_p} \xi^{i_1} \wedge \ldots \wedge \xi^{i_p}. \tag{7.60}$$

In what follows the vector ξ will be designated as $dx = (dx^1, dx^2, \ldots, dx^n)$ and the vectors ξ_h as $d_h x = (d_h x^1, d_h x^2, \ldots, d_h x^n)$. We choose as a basis in E^n the vectors $e_h = \{0, 0, \ldots, 1, 0, \ldots, 0\}$, where the unity is in the kth place. Elements of the conjugate basis are the functions $e^h(\xi) = e^k(dx)$ defined by the equations

$$e^k(dx) = dx^k.$$

Then the differential form (7.60) becomes

$$\omega(x, d_1x, \ldots, d_px) = \sum_{i_1<\ldots<i_p} \omega_{i_1\ldots i_p}(x)\, dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Example 1. A differential 0-form is any function defined in a domain G (and, by virtue of our assumptions, infinitely differentiable in G).

Example 2. A differentiable 1-form is

$$\omega(x, dx) = \sum_{h=1}^{n} \omega_h(x)\, dx^h.$$

In particular, when $n = 1$, $\omega(x, dx) = f(x)\, dx$. A differentiable form of degree 1 is also called a linear differentiable form.

Example 3. A differentiable 2-form is

$$\omega(x, d_1x, d_2x) = \sum_{i<h} \omega_{ih}(x)\, dx^i \wedge dx^h.$$

By definition

$$dx^i \wedge dx^k = (e^i \wedge e^h)(d_1x, d_2x) = e^i(d_1x)\, e^h(d_2x) - e^i(d_2x)\, e^h(d_1x) =$$
$$= d_1x^i\, d_2x^h - d_2x^i\, d_1x^h = \begin{vmatrix} d_1x^i & d_1x^h \\ d_2x^i & d_2x^h \end{vmatrix}.$$

In particular, for $n = 2$ we get

$$\omega(x, d_1x, d_2x) = f(x) \begin{vmatrix} d_1x^1 & d_1x^2 \\ d_2x^1 & d_2x^2 \end{vmatrix}.$$

The determinant is equal to the area element corresponding to the vectors d_1x and d_2x.

In the case where $n = 3$, denoting $\omega_{12} = R$, $\omega_{23} = P$, $\omega_{13} = -Q$, we get

$$\omega = P\, dx^2 \wedge dx^3 - Q\, dx^1 \wedge dx^3 + R\, dx^1 \wedge dx^2 = \begin{vmatrix} P & Q & R \\ d_1x^1 & d_1x^2 & d_1x^3 \\ d_2x^1 & d_2x^2 & d_2x^3 \end{vmatrix}.$$

Example 4. A differential 3-form in three-dimensional space is

$$\omega(x, d_1x, d_2x, d_3x) = f(x)\, dx^1 \wedge dx^2 \wedge dx^3 = f(x)\begin{vmatrix} d_1x^1 & d_1x^2 & d_1x^3 \\ d_2x^1 & d_2x^2 & d_2x^3 \\ d_3x^1 & d_3x^2 & d_3x^3 \end{vmatrix}.$$

The determinant is equal to the volume element corresponding to the vectors d_1x, d_2x, d_3x.

7S.2.2. The exterior differential.

Definition. *The exterior differential of a p-linear differential form $\omega \in \Omega_p(G)$ is a form $d\omega \in \Omega_{p+1}(G)$ defined by the relation*

$$d\omega = \sum_{i_1 < \ldots < i_p} d\omega_{i_1 \ldots i_p} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p},$$

where

$$d\omega_{i_1 \ldots i_p} = \sum_{k=1}^{n} \frac{\partial \omega_{i_1 \ldots i_p}}{\partial x^k}\, dx^k.$$

Thus, if

$$\omega = \sum_{i_1 < \ldots < i_p} \omega_{i_1 \ldots i_p}\, dx^{i_1} \wedge \ldots \wedge dx^{i_p},$$

then

$$d\omega = \sum_{k=1}^{n} \sum_{i_1 < \ldots < i_p} \frac{\partial \omega_{i_1 \ldots i_p}}{\partial x^k}\, dx^k \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Example 1. The differential of a form of zero degree (i.e. of the function $f(x)$) has the form

$$df(x) = \sum_{k=1}^{n} \frac{\partial f}{\partial x^k}\, dx^k.$$

Example 2. Evaluate the differential of the linear form

$$\omega = \omega(x, dx) = \sum_{i=1}^{n} \omega_i(x)\, dx^i.$$

We get

$$d\omega = d\omega(x, d_1x, d_2x) = \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{\partial \omega_i(x)}{\partial x^k}\, dx^k \wedge dx^i.$$

Since $dx^k \wedge dx^i = -dx^i \wedge dx^k$ and $dx^k \wedge dx^k = 0$, we have

$$\begin{aligned} d\omega &= \sum_{k<i} \frac{\partial \omega_i}{\partial x^k}\, dx^k \wedge dx^i + \sum_{i<k} \frac{\partial \omega_i}{\partial x^k}\, dx^k \wedge dx^i = \\ &= \sum_{k<i} \frac{\partial \omega_i}{\partial x^k}\, dx^k \wedge dx^i - \sum_{k<i} \frac{\partial \omega_k}{\partial x^i}\, dx^k \wedge dx^i = \\ &= \sum_{k<i} \left(\frac{\partial \omega_i}{\partial x^k} - \frac{\partial \omega_k}{\partial x^i} \right) dx^k \wedge dx^i. \end{aligned}$$

In particular, when $n = 2$ we get for $\omega = P\,dx^1 + Q\,dx^2$

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx^1 \wedge dx^2.$$

78.2.3. Properties of exterior differential. The following properties follow immediately from the definition

(1) if $\omega_1 \in \Omega_p(G)$, $\omega_2 \in \Omega_p(G)$ then $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$,
(2) if $\omega \in \Omega_p(G)$ and λ is a real number, then $d(\lambda\omega) = \lambda\, d\omega$,
(3) if $\omega_1 \in \Omega_p(G)$, $\omega_2 \in \Omega_q(G)$, then

$$d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^p \omega_1 \wedge d\omega_2.$$

We prove property (3). Let

$$\omega = \sum_{i_1 < \ldots < i_p} \omega_{i_1 \ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

We introduce the following notation:

$$\frac{\partial \omega}{\partial x^k} = \sum_{i_1 < \ldots < i_p} \frac{\partial \omega_{i_1 \ldots i_p}}{\partial x^k} dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Then $d\omega$ may be written as

$$d\omega = \sum_{k=1}^{n} dx^k \wedge \frac{\partial \omega}{\partial x^k}.$$

Remember that

$$\omega = \omega_1 \wedge \omega_2 = (-1)^{pq} \omega_2 \wedge \omega_1.$$

Further

$$\frac{\partial \omega}{\partial x^k} = \frac{\partial \omega_1}{\partial x^k} \wedge \omega_2 + \omega_1 \wedge \frac{\partial \omega_2}{\partial x^k} = \frac{\partial \omega_1}{\partial x^k} \wedge \omega_2 + (-1)^{pq} \frac{\partial \omega_2}{\partial x^k} \wedge \omega_1.$$

Then

$$d\omega = \sum_{k=1}^{n} dx^k \wedge \frac{\partial \omega}{\partial x^k} = \sum_{k=1}^{n} dx^k \wedge \frac{\partial \omega_1}{\partial x^k} \wedge \omega_2 +$$

$$+ (-1)^{pq} \sum_{k=1}^{n} dx^k \wedge \frac{\partial \omega_2}{\partial x^k} \wedge \omega_1 = d\omega_1 \wedge \omega_2 + (-1)^{pq} d\omega_2 \wedge \omega_1.$$

Since $d\omega_2$ is a $(q+1)$-form, we have

$$d\omega_2 \wedge \omega_1 = (-1)^{p(q+1)} \omega_1 \wedge d\omega_2.$$

Hence $d\omega = d\omega_1 \wedge \omega_2 + (-1)^p \omega_1 \wedge d\omega_2$.

The following *important* property of the differential is true.

The basic property of the exterior differential:

$$d(d\omega) = 0.$$

Proof. Suppose first that ω is a form of degree 0, i.e. $\omega(x) = f(x)$. Then

$$d(df) = d \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i = \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^k \partial x^i} dx^k \wedge dx^i.$$

Since $dx^k \wedge dx^i = -dx^i \wedge dx^k$, this equation may be rearranged in the form

$$d(df) = \sum_{i<k} \left(\frac{\partial^2 f}{\partial x^i \partial x^k} - \frac{\partial^2 f}{\partial x^k \partial x^i} \right) dx^i \wedge dx^k,$$

hence $d(df) = 0$.

Now let

$$\omega = \sum_{i_1<\ldots<i_p} \omega_{i_1\ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Then

$$d\omega = \sum_{k=1}^{n} \sum_{i_1<\ldots<i_p} d\omega_{i_1\ldots i_p} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}.$$

Notice that each term of the sum is the external product of the differentials of 0-forms, i.e. of forms $\omega_{i_1\ldots i_p}(x)$, $e^{i_1}(dx)$, . . ., $e^{i_p}(dx)$. It remains to apply property (3) and take advantage of the fact that for the form of degree 0 the basic property has been proved.

7S.3. Differentiable Mappings

7S.3.1. Definition of differentiable mappings. Consider an arbitrary m-dimensional domain D of a Euclidean space E^m and an n-dimensional domain $G \subset E^n$. Points of D are designated $t = (t^1, t^2, \ldots, t^m)$ and those of G are designated $x = (x^1, x^2, \ldots, x^n)$.

We shall say that φ maps D into G if

$$\varphi = \{\varphi^1, \varphi^2, \ldots, \varphi^n\},$$

where $\varphi^k(t)$ are defined in D and vectors x with coordinates $x^k = \varphi^k(t)$ lie in G.

We define a mapping φ^* transforming $\Omega_p(G)$ into $\Omega_p(D)$ for any p, $0 \leqslant p \leqslant n$. We shall assume that each component $\varphi^k(t)$ of a mapping φ is infinitely differentiable.

Definition. *Let φ be a mapping of $D \subset E^m$ into $G \subset E^n$. We denote by φ^* the mapping which for all $0 \leqslant p \leqslant n$ acts from $\Omega_p(G)$ into $\Omega_p(D)$ by the following rule: if*

$$\omega = \sum_{i_1<\ldots<i_p} \omega_{i_1\ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p},$$

then

$$\varphi^*(\omega) = \sum_{i_1<\ldots<i_p} \omega_{i_1\ldots i_p}(\varphi(t))\, \varphi^*(dx^{i_1}) \wedge \ldots \wedge \varphi^*(dx^{i_p}),$$

where

$$\varphi^*(dx^i) = \sum_{k=1}^{m} \frac{\partial \varphi^i}{\partial t^k} dt^k.$$

Example 1. Let ω be a form of degree 0, i.e. $\omega = f(x)$. Then

$$\varphi^*(f) = f(\varphi(t)).$$

Example 2. Let φ map an n-dimensional domain $D \subset E^n$ into an n-dimensional domain $G \subset E^n$, and let ω be the following n-form:

$$\omega = dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n.$$

Then

$$\varphi^*(\omega) = \left(\sum_{k_1=1}^{n} \frac{\partial \varphi^1}{\partial t^{k_1}} dt^{k_1}\right) \wedge \ldots \wedge \left(\sum_{k_n=1}^{n} \frac{\partial \varphi^n}{\partial t^{k_n}} dt^{k_n}\right) =$$

$$= \sum_{k_1=1}^{n} \sum_{k_n=1}^{n} \frac{\partial \varphi^1}{\partial t^{k_1}} \ldots \frac{\partial \varphi^n}{\partial t^{k_n}} \partial t^{k_1} \wedge \ldots \wedge dt^{k_n} =$$

$$= dt^1 \wedge \ldots \wedge dt^n \sum_{\sigma} (\operatorname{sgn} \sigma) \frac{\partial \varphi^1}{\partial t^{\sigma(1)}} \ldots \frac{\partial \varphi^n}{\partial t^{\sigma(n)}} =$$

$$= dt^1 \wedge \ldots \wedge dt^n \det \left\{\frac{\partial \varphi^i}{\partial t^j}\right\}.$$

Thus

$$\varphi^*(dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n) = \frac{D(\varphi^1, \varphi^2, \ldots, \varphi^n)}{D(t^1, t^2, \ldots, t^n)} dt^1 \wedge dt^2 \wedge \ldots \wedge dt^n.$$

Remark. The form $\varphi^*(\omega)$ is a differential form obtained from a form ω by changing variables φ.

78.3.2. Properties of mapping φ^*. The following properties of the mapping φ^* are true:

(1) If $\omega_1 \in \Omega_p(G)$, $\omega_2 \in \Omega_q(G)$, then

$\varphi^*(\omega_1 \wedge \omega_2) = \varphi^*(\omega_1) \wedge \varphi^*(\omega_2)$.

Proof. Let

$$\omega_1 = \sum_{i_1 < \ldots < i_p} a_{i_1 \ldots i_p}(x)\, dx^{i_1} \wedge \ldots \wedge dx^{i_p},$$

$$\omega_2 = \sum_{k_1 < \ldots < k_q} b_{k_1 \ldots k_q}(x)\, dx^{k_1} \wedge \ldots \wedge dx^{k_q}.$$

Then

$$\omega_1 \wedge \omega_2 = \sum_{i_1 < \ldots < i_p} \sum_{k_1 < \ldots < k_q} a_{i_1, \ldots i_p}(x)\, b_{k_1, \ldots k_q}(x) \times$$

$$\times dx^{i_1} \wedge \ldots \wedge dx^{i_p} \wedge dx^{k_1} \wedge \ldots \wedge dx^{k_q}$$

and hence

$$\varphi^*(\omega_1 \wedge \omega_2) = \sum_i \sum_k a_i(\varphi(t))\, b_k(\varphi(t))\, \varphi^*(dx^{i_1}) \wedge \ldots \wedge \varphi^*(dx^{k_q}) =$$

$$= \sum_i a_i(\varphi)\, \varphi^*(dx^{i_1}) \wedge \ldots \wedge \varphi^*(dx^{i_p}) \wedge \left[\sum_k b_k(\varphi)\, \varphi^*(dx^{k_1}) \wedge \ldots\right.$$

$$\left.\ldots \wedge \varphi^*(dx^{k_q})\right] = \varphi^*(\omega_1) \wedge \varphi^*(\omega_2).$$

(2) If $\omega \in \Omega_p(G)$, then

$\varphi^*(d\omega) = d\varphi^*(\omega)$.

Proof. We first prove this equation for $p = 0$, i.e. for $\omega = f(x)$. We get

$$d\omega = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx^i, \quad \varphi^*(\omega) = f(\varphi(t)),$$

$$d\varphi^*(\omega) = \sum_{h=1}^{m} \frac{\partial}{\partial t^h} f(\varphi(t))\, dt^h = \sum_{h=1}^{m} \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \frac{\partial \varphi^i}{\partial t^h} dt^h =$$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \varphi^*(dx^i) = \varphi^*(d\omega).$$

For an arbitrary p, we carry out a proof by induction. Let $\omega = f_{i_1 \ldots i_p}(x)\, dx^{i_1} \wedge \wedge \ldots \wedge dx^{i_p}$. Then $d\omega = df_{i_1 \ldots i_p} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}$. By property (1) and the just proved relation

$$\varphi^*(d\omega) = \varphi^*(df) \wedge \varphi^*(dx^{i_1}) \wedge \ldots \wedge \varphi^*(dx^{i_p}).$$

On the other hand,

$$d\varphi^*(\omega) = d\varphi^*[(f\, dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}) \wedge dx^{i_p}] =$$

$$= d[\varphi^*(f\, dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}) \wedge \varphi^*(dx^{i_p})].$$

Further, by property (3) of the exterior differential

$$d\varphi^*(\omega) = d\varphi^*(f\, dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}) \wedge \varphi^*(dx^{i_p}) +$$

$$+ (-1)^{p-1} \varphi^*(f\, dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}) \wedge d\varphi^*(dx^{i_p}).$$

Notice that $\varphi^*(dx^{i_p}) = d\varphi^*(x^{i_p})$ by what has just been proved, and then by a basic property of the exterior differential $d\varphi^*(dx^{i_p}) = 0$.

Under the induction hypothesis true for $p - 1$

$$d\varphi^*(f\, dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}) = \varphi^*(df \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}).$$

As a result we get

$$d\varphi^*(\omega) = \varphi^*(df \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}) \wedge \varphi^*(dx^{i_p})$$

and by property (1)

$$d\varphi^*(\omega) = \varphi^*(df \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}).$$

The next important property is *transitivity*.

(3) Consider open domains $U \subset E^l$, $V \subset E^m$, $W \subset E^n$ whose points are $u = (u^1, u^2, \ldots, u^l)$, $v = (v^1, v^2, \ldots, v^m)$, $w = (w^1, w^2, \ldots, w^n)$, respectively. Let φ map $U \to V$ and let ψ map $V \to W$. By $\psi \circ \varphi$ denote a mapping, called composition, which acts by the rule

$$(\psi \circ \varphi)(u) = \psi[\varphi(u)].$$

Similarly we introduce the composition $\varphi^* \circ \psi^*$ which for any p transforms $\Omega_p(W)$ into $\Omega_p(U)$, i.e.

$$(\varphi^* \circ \psi^*)(\omega) = \varphi^*[\psi^*(\omega)].$$

The following statement is true: *if $C_1 = C_2$, then*

$$\int_{C_1} \omega = \int_{C_2} \omega.$$

Proof. We give a proof for the case where

$\omega = f(x)\, dx^1 \wedge dx^2 \wedge \ldots \wedge dx^p.$

By definition

$$\int_{C_2} \omega = \int_{I^p} f[\varphi_2(t)] \frac{D(\varphi_2^1, \varphi_2^2, \ldots, \varphi_2^p)}{D(t^1, \ldots, t^2, \ldots, t^p)} dt^1 \wedge \ldots \wedge dt^p.$$

As stated there is a mapping τ of a cube I^p onto itself satisfying conditions (1) and (2).

We make in the integral a change of the variable $t = \tau(s)$, $s \in I^p$. We get $\varphi_2(t) = \varphi_2[\tau(s)] = \varphi_1(s)$,

$$\int_{C_2} \omega = \int_{I^p} f[\varphi_1(s)] \frac{D(\varphi_2^1, \varphi_2^2, \ldots, \varphi_2^p)\, D(\tau^1, \tau^2, \ldots, \tau^p)}{D(t^1, t^2, \ldots, t^p)\, D(s^1, s^2, \ldots, s^p)} ds^1 \wedge ds^2 \wedge \ldots \wedge ds^p =$$

$$= \int_{I^p} f[\varphi_1(s)] \frac{D(\varphi_2^1, \ldots, \varphi_2^p)}{D(s^1, \ldots, s^p)} ds^1 \wedge \ldots \wedge ds^p = \int_{C_1} \omega.$$

Similarly we could show that if $C_1 = -C_2$, then

$$\int_{C_1} \omega = -\int_{C_2} \omega.$$

78.4.2. Differentiable chains. We shall need surfaces that split into several pieces, each being the image of some m-dimensional cube. The two circles of the boundary of a two-dimensional ring may serve as an example of such a surface. We shall distinguish between the orientations of these circles. It appears very useful in this connection to introduce linear combinations of singular cubes with real coefficients.

Definition 1. A p-dimensional chain C is an arbitrary set

$\{\lambda_1, \lambda_2, \ldots, \lambda_h, C_1, C_2, \ldots, C_h\}$,

where λ_i are real numbers and C_i are p-dimensional singular cubes. We shall use the notation

$$C = \lambda_1 C_1 + \ldots + \lambda_h C_h.$$

We shall say that C is in G if all C_i are in G.

The set of p-dimensional chains will form a linear space if we introduce in a natural way the operations of addition and multiplication by real numbers.

Definition 2. The integral of a form ω over a p-dimensional chain C contained in G is

$$\int_C \omega = \lambda_1 \int_{C_1} \omega + \lambda_2 \int_{C_2} \omega + \ldots + \lambda_h \int_{C_h} \omega.$$

Now we can define the boundary of an arbitrary singular cube. To do this we first define the boundary of a unit cube.

Definition 3. The boundary of a cube I^p is a $(p-1)$-dimensional chain

$$\partial I^p = \sum_{i=1}^{p} (-1)^i \left[I_0^p(i) - I_1^p(i)\right],$$

where $I_\alpha^p(i)$ is the intersection of the cube I^p with the hyperplane $x^i = \alpha$, $(\alpha = 0, 1)$.

For this definition to be correct, it is necessary to make clear how we interpreted the statement that $I_\alpha^p(i)$ is a $(p-1)$-dimensional singular cube.

We construct the *canonical* mapping $\tilde{\varphi} = \tilde{\varphi}_i^{\alpha,p}$ of a cube I^{p-1} on $I_\alpha^p(i)$. Let $s = (s^1, s^2, \ldots, s^{p-1}) \in I^{p-1}$. We set

$$\tilde{\varphi}^k(s) = \begin{cases} s^k & \text{if } 1 \leqslant k < i, \\ \alpha & \text{if } k = i, \\ s^{k-1} & \text{if } i < k \leqslant p. \end{cases}$$

Obviously $\tilde{\varphi} = (\tilde{\varphi}^1, \tilde{\varphi}^2, \ldots, \tilde{\varphi}^p)$ maps I^{p-1} in a one-to-one manner onto $I_\alpha^p(i)$. For $\alpha = 0$ and $i = p$ in particular the mapping φ is the restriction to $I_0^p(p-1)$ of the identity mapping of a space E^p onto itself.

Definition 4. The boundary of a p-dimensional singular cube $C = \varphi\colon I^p \to E^n$ is a $(p-1)$-dimensional chain

$$\partial C = \sum_{i=1}^{p} (-1)^i \left[\varphi(I_0^p(i)) - \varphi(I_1^p(i))\right].$$

The boundary of the image of I^p is thus the image of the boundary of I^p with natural orientation.

Example 1. Consider in the plane a square I^2. Obviously this square may be considered as a singular cube if we take as φ an identity mapping. Figure 7.11

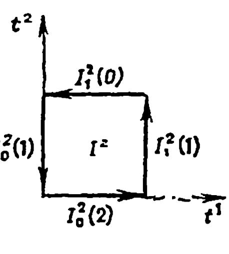


Fig. 7.11

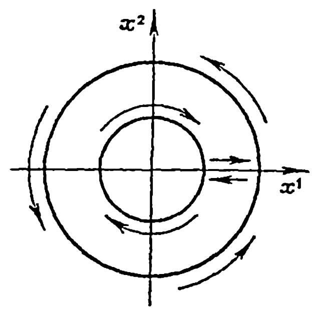


Fig. 7.12

shows the boundary of the square. Here the direction of the arrows coincides with the direction of increase in the parameter t_k with respect to which the integration is performed, if this side of the square appears in the chain ∂I^2 with the plus sign, and is opposite if the side of the square is taken with the minus sign. We see that our convention about signs leads to the usual counterclockwise tracing of the boundary.

Example 2. Consider the singular cube $C = \varphi\colon I^2 \to R^2$, where φ has the orm

$$\varphi^1 = (a + Rt^1)\cos 2\pi t^2,$$
$$\varphi^2 = (a + Rt^1)\sin 2\pi t^2.$$

It is easy to see that $\varphi(I^2)$ is a ring whose boundary is formed by circles of radii a and $a+R$. We shall show what the boundary of a singular cube C is. Obviously $\varphi(I_0^2(1))$ is the circle

$$\varphi^1 = a\cos 2\pi t^2,$$
$$\varphi^2 = a\sin 2\pi t^2.$$

Further, $\varphi(I_1^2(1))$ is a circle of radius $a+R$. Finally, $\varphi(I_0^2(2))$ and $\varphi(I_1^2(2))$ is a segment $x^2=0$, $a \leqslant x^1 \leqslant a+R$.

In Fig. 7.12 the arrows indicate the direction of circulation about the boundary ∂C if the boundary ∂I^2 is traced counterclockwise.

Since $\varphi(I_0^2(2)) - \varphi(I_1^2(2)) = 0$, we may assume that

$$\partial C = \varphi(I_1^2(1)) - \varphi(I_0^2(1)),$$

which coincides with the usual interpretation of the boundary of a ring.

We shall see in what way the integral of a form ω along the boundary of a cube C and that of a form $\varphi^*(\omega)$ along the boundary of I^p are related.

Statement. Let $C=\varphi\colon I^p \to E^n$ be an arbitrary singular cube contained in G and let $\omega \in \Omega_{p-1}(G)$. We have

$$\int\limits_{\partial C} \omega = \int\limits_{\partial I^p} \varphi^*(\omega).$$

Proof. Obviously, by the definition of the integral over a chain it suffices to prove the equation

$$\int\limits_{\varphi(I_\alpha^p(i))} \omega = \int\limits_{I_\alpha^p(i)} \varphi^*(\omega).$$

Consider the canonical mapping $\tilde{\varphi} = \tilde{\varphi}_i^{\alpha,p}\colon I^{p-1} \to I_\alpha^p(i)$. By definition

$$\int\limits_{I_\alpha^p(i)} \varphi^*(\omega) = \int\limits_{I^{p-1}} \tilde{\varphi}^*[\varphi^*(\omega)].$$

By property (3) of differentiable mappings (see Section 7S.3.2)

$$\tilde{\varphi}^* \circ \varphi^* = (\varphi \circ \tilde{\varphi})^*.$$

Thus

$$\int\limits_{I_\alpha^p(i)} \varphi^*(\omega) = \int\limits_{I^{p-1}} (\varphi\circ\tilde{\varphi})^*(\omega) = \int\limits_{(\varphi\circ\tilde{\varphi})(I^{p-1})} \omega = \int\limits_{\varphi(I_\alpha^p(i))} \omega,$$

since $(\varphi\circ\tilde{\varphi})(I^{p-1}) = \varphi(I_\alpha^p(i))$.

7.S.4.3. The Stokes formula.

Main theorem. Let $C=\varphi\colon I^p \to E^n$ be an arbitrary singular cube contained in G and let $\omega \in \Omega_{p-1}(G)$. We have the Stokes formula

$$\int\limits_C d\omega = \int\limits_{\partial C} \omega.$$

We prove the Stokes formula first for the following special case.

Let ω be a differential form of degree $p-1$ defined in I^p. Then

$$\int\limits_{I^p} d\omega = \int\limits_{\partial I^p} \omega. \tag{7.61}$$

Proof. Let $\omega = f(t)\, dt^2 \wedge \ldots \wedge dt^p$. By definition

$$\int_{\partial I^p} \omega = \sum_{i=1}^{p} (-1)^i \Big(\int_{I_0^p(i)} \omega - \int_{I_1^p(i)} \omega \Big).$$

Evaluate the following integral:

$$\int_{I_\alpha^p(i)} \omega, \text{ where } i=1, 2, \ldots, p, \quad \alpha=0, 1.$$

Consider the canonical mapping $\tilde{\varphi}$: $I^{p-1} \to I_\alpha^p(i)$. By virtue of the results of Section 7S.4.1

$$\int_{I_\alpha^p(i)} \omega = \int_{I^{p-1}} f[\tilde{\varphi}(s)] \frac{D(\tilde{\varphi}^2, \ldots, \tilde{\varphi}^p)}{D(s^1, \ldots, s^{p-1})} ds^1 \wedge \ldots \wedge ds^{p-1}.$$

By the definition of the canonical mapping $\tilde{\varphi}_i^{\alpha, p}$ the Jacobian has the form

$$J = \frac{D(s^2, \ldots, s^{i-1}, \alpha, s^i, \ldots, s^{p-1})}{D(s^1, s^2, \ldots, s^{p-1})} = 0, \text{ if } i \neq 1$$

and

$$J = \frac{D(s^1, s^2, \ldots, s^{p-1})}{D(s^1, s^2, \ldots, s^{p-1})} = 1 \text{ if } i=1.$$

Thus only the integrals over $I_\alpha^p(1)$ may be nonzero and we get

$$\int_{\partial I^p} \omega = (-1) \Big(\int_{I_0^p(1)} \omega - \int_{I_1^p(1)} \omega \Big) =$$

$$= \int_{I^{p-1}} f(1, s^1, s^2, \ldots, s^{p-1})\, ds^1 \wedge \ldots \wedge ds^{p-1} -$$

$$- \int_{I^{p-1}} f(0, s^1, \ldots, s^{p-1})\, ds^1 \wedge \ldots \wedge ds^{p-1}.$$

By the definition of the integral over the cube I^{p-1}

$$\int_{\partial I^p} \omega = \int_0^1 \ldots \int_0^1 [f(1, s^1, \ldots, s^{p-1}) - f(0, s^1, \ldots, s^{p-1})]\, ds^1\, ds^2 \ldots ds^{p-1} =$$

$$= \int_0^1 \int_0^1 \ldots \int_0^1 \frac{\partial f}{\partial s^0} ds^0\, ds^1 \ldots ds^{p-1} = \int_{I^p} \frac{\partial f}{\partial s^0} ds^0 \wedge \ldots \wedge ds^{p-1}.$$

On the other hand,

$$d\omega = \frac{\partial f}{\partial t^1} dt^1 \wedge dt^2 \wedge \ldots \wedge dt^p.$$

Therefore

$$\int_{I^p} d\omega = \int_{I^p} \frac{\partial f}{\partial t^1} dt^1 \wedge \ldots \wedge dt^p.$$

Thus equation (7.61) is proved.

Proof of the Stokes theorem. By the definition of the integral over the singular cube

$$\int_C d\omega = \int_{I^p} \varphi^* (d\omega).$$

By property (2) of differentiable mappings (see Section 7S.3.2)

$$\int_{I^p} \varphi^* (d\omega) = \int_{I^p} d\varphi^* (\omega).$$

Next we use the already proved Stokes formula for the cube I^p

$$\int_{I^p} d\varphi^* (\omega) = \int_{\partial I^p} \varphi^* (\omega).$$

It remains to notice that by the property of the integrals along the boundary of a singular cube (see this at the end of Section 7S.4.2)

$$\int_{\partial I^p} \varphi^* (\omega) = \int_{\partial C} \omega.$$

This completes the proof of the theorem.

7S.4.4. Examples. (1) Consider the case $p = 1$. A one-dimensional singular cube C in E^n is some curve whose end points are denoted by a and b. The Stokes formula takes the form

$$\int_C df = \int_{\partial C} f = f(b) - f(a).$$

In particular, when $n = 1$ we obtain the Newton-Leibniz formula

$$\int_a^b f'(x)\, dx = f(b) - f(a).$$

(2) Now let $p = 2$. A two-dimensional singular cube C is a two-dimensional surface, and a form $\omega \in \Omega_1$ is

$$\omega = \sum_{h=1}^{n} \omega_h\, dx^h.$$

Using Example 2 of Section 7S.2.2 we get

$$\int_C \sum_{k<l} \left(\frac{\partial \omega^l}{\partial x^k} - \frac{\partial \omega^k}{\partial x^l} \right) dx^k \wedge dx^l = \int_{\partial C} \sum_{h=1}^{n} \omega_h\, dx^h.$$

If $n = 2$, then, denoting $\omega = P\,dx^1 + Q\,dx^2$, we obtain the Green formula

$$\int\limits_C \left(\frac{\partial Q}{\partial x^1} - \frac{\partial P}{\partial x^2}\right) dx^1 \wedge dx^2 = \int\limits_{\partial C} P\,dx^1 + Q\,dx^2.$$

If $n = 3$, we obtain the ordinary Stokes formula.

(3) Let $p = n$. Then $\omega \in \Omega_{n-1}$ has the form

$$\omega = \sum_{k=1}^{n} \omega_k\, dx^1 \wedge \ldots \wedge dx^{k-1} \wedge dx^{k+1} \wedge \ldots \wedge dx^n.$$

Further

$$d\omega = \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{\partial \omega_k}{\partial x^i}\, dx^i \wedge dx^1 \wedge \ldots \wedge dx^n =$$

$$= \sum_{k=1}^{n} (-1)^{k-1} \frac{\partial \omega^k}{\partial x^k}\, dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n.$$

In particular, for $n = 3$

$$\omega = P\,dx^2 \wedge dx^3 - Q\,dx^1 \wedge dx^3 + R\,dx^1 \wedge dx^2,$$

$$d\omega = \left(\frac{\partial P}{\partial x^1} + \frac{\partial Q}{\partial x^2} + \frac{\partial R}{\partial x^3}\right) dx^1 \wedge dx^2 \wedge dx^3,$$

and we obtain the Ostrogradsky formula.

CHAPTER 8

THE LEBESGUE INTEGRAL AND MEASURE

In Chapter 10 of Volume 1 and in Chapter 2 of this volume we studied the Riemann integral of a function of one and n variables respectively. The concept of Riemann integral covered a class of functions that are either strictly continuous in a domain under consideration or close to continuous functions (the set of whose points of discontinuity has an n-dimensional volume equal to zero). This notion turns out to be insufficient in a number of fundamental branches of modern mathematics (in the theory of generalized functions, in the modern theory of partial differential equations and so on).

In this chapter we present a theory of a more general integral, the so-called *Lebesgue integral**, for which purpose we develop as a preliminary a theory of measure and the so-called *measurable functions* (which are a broad generalization of continuous functions).

The basic idea of the Lebesgue integral distinguishing it from the Riemann integral is that in composing a Lebesgue integral sum the points are combined into individual terms not by the principle of their closeness in the domain of integration (as was in the Riemann integral sum) but by the principle of the closeness of the values of the integrable function at those points. It is this idea that allows the notion of integral to be extended to include a very wide class of functions.

It is to be noted that many mathematical theories allowing interpretation of the integral in the Riemann sense assume a more complete character when the Lebesgue integral is used. An example of such a theory is the Fourier series theory presented using the integral in the Riemann sense in Chapter 10 and the Lebesgue integral in Chapter 11.

Throughout this chapter the presentation is made for the case of a single variable but can be extended without difficulty to any number n of variables (an appropriate remark is made at the end of the chapter).

* Henri Lebesgue (1875-1941), a French mathematician.

8.1. ON THE STRUCTURE OF OPEN AND CLOSED SETS

We shall consider an arbitrary set E of points of an infinite straight line $(-\infty, \infty)$.

The *complement* of E is the set designated CE and equal to the collection of those points of the infinite straight line $(-\infty, \infty)$ which are not in E.

If we regard as the *difference* of sets A and B the collection of those points of A which are not in B and denote the difference of A and B by $A \setminus B$, then the complement CE of E may be represented as

$$CE = (-\infty, \infty) \setminus E.$$

We recall some definitions introduced in Volume 1.

1°. A point x is said to be an *interior point* of a set E if there is some neighbourhood of x (i.e. an interval containing that point) lying entirely in E.

In what follows an arbitrary neighbourhood of x will be designated $v(x)$.

2°. A point x is said to be a *limit point* of E if in any neighbourhood $v(x)$ of x there is at least one point x^1 of E distinct from x.

3°. A set G is said to be *open* if all the points of that set are interior.

4°. A set F is said to be *closed* if it contains all of its limit points*.

Let us agree to denote the collection of all limit points of an arbitrary set E by E' and the *sum* or *union* of two sets A and B by $A+B$ or $A \cup B$**. Let us also agree to give the name of *closure* of an arbitrary set E to a set designated $\overline{E}$ and equal to the sum $E + E'$.

It is obvious that for any closed set F we have $\overline{F} = F$.

The collection of all interior points of an arbitrary set E will be designated int E***.

It is obvious that for any open set G we have $\operatorname{int} G = G$.

For an entirely arbitrary set E the set $\operatorname{int} E$ is open and $\overline{E}$ is closed.

Remark. One can show that int E is the sum of all open sets contained in E and that $\overline{E}$ is the intersection**** of all closed sets containing E. Thus int E is the *largest* open set contained in E and $\overline{E}$ is the *smallest* closed set containing E.

We shall discuss the simplest properties of open and closed sets.

1°. *If a set F is closed, then its complement CF is open.*

---

* In particular, a set having no limit points is closed (for an empty set is contained in any set).

** The *sum* or *union* of sets A and B is a set C consisting of points lying in at least one of the sets A or B.

*** int are the first three letters of the French word intérieur (interior).

**** The *intersection* of A and B is a set of points lying both in A and B.

Proof. No point x of the set CF is in F, nor is it (by the closure of F) in the set F' of the limit points of F. But this means that some neighbourhood $v(x)$ of x is not in F and is therefore in CF.

2°. *If a set G is open, then its complement CG is closed.*

Proof. Any limit point x of the set CG is *a fortiori* in that set, for otherwise x would be in G and since G is an open set some neighbourhood $v(x)$ of x would also be in G and would not in CG, i.e. x would not be a limit point of CG.

3°. *The sum of any number of open sets is an open set.*

Proof. Let a set E be the sum of any number of open sets G_α (the subscript α is not an integer in general) and let x be an arbitrary point of E. Then (by the definition of a sum of sets) x is in at least one of the sets G_α and since every set G_α is open, there is some neighbourhood $v(x)$ of x also lying in the set G_α and hence in the set E.

4°. *The intersection of any finite number of open sets is an open set.*

Proof. Let a set E be the intersection of open sets $G_1, G_2, \ldots, G_n$ and let x be any point of E. Then for any k ($k = 1, 2, \ldots, n$) the point x is in G_k and therefore there is some neighbourhood $v_k(x) = (x - \varepsilon_k, x + \varepsilon_k)$, $\varepsilon_k > 0$, of x also lying in G_k. If $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$, then the neighbourhood $v(x) = (x - \varepsilon, x + \varepsilon)$ of x is in all G_k and consequently in E.

5°. *The intersection of any number of closed sets is a closed set.*

Proof. Let a set E be the intersection of any number of closed sets F_α (the subscript α is not an integer in general). Notice that the complement CE is the sum of all complements CF_α, each of which is, according to 1°, an open set.

According to 3° the set CE is open and therefore in view of 2° the set E is closed.

6°. *The sum of a finite number of closed sets is a closed set.*

Proof. Let E be the sum of closed sets $F_1, F_2, \ldots, F_n$. Then CE is the intersection of the sets $CF_1, CF_2, \ldots, CF_n$, each of which is by 1° an open set. According to 4° the set CE is open and therefore in view of 2° the set E is closed.

7°. *If a set F is closed and a set G is open, then the set $F \setminus G$ is closed and the set $G \setminus F$ is open.*

Proof. It suffices to notice that $F \setminus G$ is the intersection of the closed sets F and CG and $G \setminus F$ is the intersection of the open sets G and CF.

Using the above properties we prove a theorem on the structure of an arbitrary open set of points of an infinite straight line.

Let us agree throughout the remainder of this chapter to apply the term interval to any connected open set (not necessarily bounded) of points of an infinite straight line. In other words, an interval is either an open interval $a < x < b$ or one of the open half-lines $a < x < \infty$ or $-\infty < x < b$ or an entire infinite straight line $-\infty < x < \infty$.

Theorem 8.1. Any open set of points of an infinite straight line is a sum of a finite or countable number of mutually disjoint intervals.*

Proof. Let G be an open set and let x be an arbitrary fixed point of G. Since G is open, there is some neighbourhood $v(x)$ of x contained in G. The sum of all neighbourhoods $v(x)$ of the given fixed point x that are contained in G will be designated $I(x)$. We prove that $I(x)$ is an *interval*.

Denote by a the infimum of the set of all points of $I(x)$ (in case the set of all points of $I(x)$ is not bounded below, we set $a = -\infty$) and by b the supremum of the set of all points of $I(x)$ (in case the set of all points of $I(x)$ is not bounded above, we set $b = \infty$). It is sufficient to prove that an arbitrary point y of the interval (a, b) is in $I(x)$. Let y be an arbitrary point of (a, b). To be specific assume that $a < y < x$) (the case $x < y < b$ is considered quite similarly). By the definition of infimum there is a point y' in $I(x)$ such that $a \leqslant y' < y$. But this means that there is some neighbourhood $v(x)$ of the point x we have fixed containing the point y'. By virtue of the inequality $y' < y < x$ that same neighbourhood $v(x)$ contains the point y. It follows that $I(x)$ too contains y and that the proof that $I(x)$ is an interval is complete. One may say that $I(x)$ is the *largest* interval containing the point x and contained in G.

We now show that if interval $I(x_1)$ and $I(x_2)$ are constructed for two distinct fixed points x_1 and x_2 of G then those intervals *either have no points in common or coincide*. Indeed, if $I(x_1)$ and $I(x_2)$ contained a common point x, then they would be both contained in $I(x)$ and would therefore coincide.

Having constructed for each of the points x an interval $I(x)$ we now select the intervals containing no points in common (i.e. the ones that are mutually disjoint). Such intervals each contained at least one rational point (we know this from Chapter 2 of [1]). Since the set of all rational points is countable (see Section 3.4.6 of [1]), the number of all mutually disjoint intervals $I(x)$ is at most countable. Since the sum of all such intervals is a set G, the theorem is proved.

Corollary. Any closed set of points of an infinite straight line is obtained by removing from the line a finite or countable number of mutually disjoint intervals.

8.2. MEASURABLE SETS

8.2.1. The outer measure of a set and its properties. The entire theory presented in this section is due to H. Lebesgue. Its starting point is using as a primary (original) set the interval $\Delta = (a, b)$

* Recall that a *countable* set is an infinite set whose elements can be enumerated, i.e. brought to correspond in a one-to-one manner to the natural number series 1, 2, 3, . . . (see Section 3.4.6 in [1]).

whose length or measure is assumed to be known and equal to a number $|\Delta| = b - a > 0$.

Let E be an arbitrary set on the number line. A *covering* $S = S(E)$ of E is any finite or countable system of intervals $\{\Delta_n\}$ whose sum contains the set E. The sum of the lengths of all the intervals $\{\Delta_n\}$ constituting a covering $S = S(E)$ is designated $\sigma(S)$. So

$$\sigma(S) = \sum_n |\Delta_n| \leqslant \infty.$$

Definition. *The outer measure of a set E is the infimum of $\sigma(S)$ on the set of all coverings $S = S(E)$ of E.*

The outer measure of a set E will be designated $|E|^*$. So by definition

$$|E|^* = \inf_{S(E)} \sigma(S).$$

Obviously the outer measure of any interval coincides with its length.

We now show the basic properties of the outer measure.

1°. *If a set E_1 is contained in E_2^*, then $|E_1|^* \leqslant |E_2|^*$.*

To prove this it suffices to notice that any covering of E_2 is at the same time a covering of E_1.

2°. *If E is the sum of a finite or countable number of sets $\{E_k\}$ (in symbols $E = \bigcup\limits_{k=1}^{\infty} E_k$), then*

$$|E|^* \leqslant \sum_{k=1}^{\infty} |E_n|^*. \tag{8.1}$$

Proof. Fix an arbitrary $\varepsilon > 0$. By the definition of the measure $|E_k|^*$ as the infimum, for every k there is a covering $S_k(E_k)$ of a set E_l by a system of intervals $\{\Delta_n^k\}$ $(n = 1, 2, \ldots)$ such that

$$\sum_{n=1}^{\infty} |\Delta_n^k| \leqslant |E_l|^* + \frac{\varepsilon}{2^k}. \tag{8.2}$$

Denote by S the covering of the whole of E combining all the coverings S_k $(k = 1, 2, \ldots)$ and consisting of all the intervals $\{\Delta_n^k\}$ $(k = 1, 2, \ldots; n = 1, 2, \ldots)$. Since S is a covering of E, we have $|E|^* \leqslant \sigma(S)$, but $\sigma(S) = \sum\limits_{k=1}^{\infty} \sum\limits_{n=1}^{\infty} |\Delta_n^k|$.

* In symbols the fact that a set E_1 is contained in E_2 is designated as follows: $E_1 \subset E_2$.

From the last two relations and from (8.2) we get

$$|E|^* \leqslant \sum_{k=1}^{\infty} \left(|E_k|^* + \frac{\varepsilon}{2^k} \right) = \sum_{k=1}^{\infty} |E_k|^* + \varepsilon.$$

Thus inequality (8.1) is proved.

Let us agree to regard as the distance between sets E_1 and E_2 the infimum of the distances between two points of E_1 and E_2 respectively.

We shall denote the distance between E_1 and E_2 by $\rho(E_1, E_2)$.

3°. *If* $\rho(E_1, E_2) > 0$, *then* $|E_1 \cup E_2|^* = |E_1|^* + |E_2|^*$.

Proof. Set $\delta = (1/2)\rho(E_1, E_2)$. For an arbitrary $\varepsilon > 0$ and a chosen $\delta > 0$ there is a covering $S(E)$ of $E = E_1 \cup E_2$ such that $\sigma(S) \leqslant |E|^* + \varepsilon$ and the length of each interval of the covering $|\Delta_n|$ is less than δ*. It is obvious that the intervals Δ_n covering the points of E_1 contain no points of E_2 and conversely the intervals covering the points of E_2 contain no points of E_1. In other words, the covering $S(E)$ falls into the sum of two coverings $S(E) = S_1(E_1) + S_2(E_2)$, the first, S_1, covering E_1 and the second, S_2, covering E_2. So we get

$$S_1(E_1) + S_2(E_2) \leqslant |E|^* + \varepsilon.$$

It follows that $|E_1|^* + |E_2|^* \leqslant |E|^* + \varepsilon$ and therefore (by the arbitrariness of ε) $|E_1|^* + |E_2|^* \leqslant |E|^*$. Since on the basis of property 2° the opposite inequality $|E|^* \leqslant |E_1|^* + |E_2|^*$ is also true, we have $|E|^* = |E_1|^* + |E_2|^*$. Thus property 3° is proved.

In particular property 3° is true if E_1 and E_2 are bounded, closed and contain no points in common.

4°. *For an arbitrary set E and an arbitrary number $\varepsilon > 0$ there is an open set G containing E such that* $|G|^* \leqslant |E|^* + \varepsilon$.

Proof. It suffices to take as G the sum of all the intervals constituting a covering $S(E)$ of E for which $\sigma(S) \leqslant |E|^* + \varepsilon$.

8.2.2. Measurable sets and their properties.

Definition 1. A set E is said to be measurable if for any positive number ε there is an open set G containing E and such that the outer measure of the difference $G \setminus E$ is less than ε.

The outer measure of a measurable set E will be called the measure of that set and designated $|E|$.

It follows from this definition that the measure of a set E is zero if and only if the outer measure of that set is also zero.

* This follows from the fact that given *arbitrary* $\varepsilon > 0$ and $\delta > 0$ there is a covering $S(E)$ of E such that $\sigma(S) < |E|^* + \varepsilon$ and $|\Delta_n| < \delta$ (for each Δ_n of S). To see this it suffices to take a covering S' for which $\sigma(S') < |E|^* + \varepsilon/2$, divide each interval of S' into intervals of length smaller than δ and cover the ends of these last intervals with intervals the total sum of whose lengths is less than $\varepsilon/2$.

We prove a number of statements revealing the basic properties of measurable sets.

Theorem 8.2. Any open set is measurable, its measure equalling the sum of the lengths of the mutually disjoint constituent intervals.

The proof is trivial (it suffices to take $G = E$ in the definition and notice that the infimum of $\sigma(S)$ is attained on the covering S coinciding with the subdivision of E into the sum of mutually disjoint intervals).

Theorem 8.3. A sum of a finite or countable number of measurable sets is a measurable set.

Proof. Let $E = \bigcup_{n=1}^{\infty} E_n$ such that every E_n being measurable. Fix an arbitrary $\varepsilon > 0$. For every set E_n there is an open set G_n containing it such that

$$|G_n \setminus E_n|^* < \varepsilon \cdot 2^{-n}. \tag{8.3}$$

On setting $G = \bigcup_{n=1}^{\infty} G_n$ we notice that the set E is contained in G and that the difference $G \setminus E$ is contained in the sum $\bigcup_{n=1}^{\infty} (G_n \setminus E_n)$. But then from Property 2° of the outer measure (see Section 8.2.1) and from inequality (8.3) we get

$$|G \setminus E|^* \leqslant \sum_{n=1}^{\infty} |G_n \setminus E_n|^* < \varepsilon \sum_{n=1}^{\infty} 2^{-n} = \varepsilon.$$

Thus the theorem is proved.

Theorem 8.4. Any closed set F is measurable.

Proof. We prove the theorem in two steps.

1°. First suppose that the set F *is bounded.* Fix an arbitrary $\varepsilon > 0$. According to Property 4° of the outer measure (see Section 8.2.1) there is an open set G containing F and such that

$$|G|^* \leqslant |F|^* + \varepsilon. \tag{8.4}$$

By Property 7° of Section 8.1 the set $G \setminus F$ is open. By Theorem 8.1 therefore $G \setminus F$ can be represented as the sum $G \setminus F = \bigcup_{n=1}^{\infty} \Delta_n$ of mutually disjoint intervals Δ_n. The theorem will be proved if we establish that

$$|G \setminus F|^* = \sum_{n=1}^{\infty} |\Delta_n| \leqslant \varepsilon. \tag{8.5}$$

For every interval $\Delta = (a, b)$ and for every number α in the interval $0 < \alpha < (b-a)/2$, let us agree to denote by Δ^α the interval $\Delta^\alpha = (a + \alpha,\ b - \alpha)$ and by $\bar{\Delta}^\alpha$ the closed interval $\bar{\Delta}^\alpha =$

$=[a+\alpha,\ b-\alpha]$. If, however, $\alpha \geqslant (b-a)/2$, then Δ^α will denote an empty set for which $|\Delta^\alpha|=0$. For every n set $\overline{E}_n^\alpha = \bigcup_{k=1}^{n} \Delta_k^\alpha$. It is obvious that $|\overline{E}_n^\alpha|^* = \sum_{k=1}^{n} |\Delta_k^\alpha|$. According to Property 6° of Section 8.1 a set $\overline{E}_n^\alpha$ is closed. Since it has no points in common with the closed set F, we have (by Property 3° of the *outer measure*)

$$|\overline{E}_n^\alpha + F|^* = |\overline{E}_n^\alpha|^* + |F|^*. \tag{8.6}$$

On the other hand, since the set $\overline{E}_n^\alpha + F$ (for any $\alpha > 0$ and for every n) is contained in G, we have (by Property 1° of the outer measure)

$$|\overline{E}_n^\alpha + F|^* \leqslant |G|^*. \tag{8.7}$$

From (8.4), (8.6), and (8.7) we get

$$|\overline{E}_n^\alpha|^* + |F|^* \leqslant |F|^* + \varepsilon \tag{8.8}$$

(for all $\alpha > 0$ and all n). Since F is bounded and its outer measure $|F|^* < \infty$, from (8.8) we get

$$|\overline{E}_n^\alpha|^* < \varepsilon \tag{8.9}$$

(for all $\alpha > 0$ and all n). Proceeding in (8.9) to the limit, first for $\alpha \to 0+0$ and then for $n \to \infty$, we obtain inequality (8.5). This proves the theorem for the case of the bounded set F.

2°. If the set F is not bounded in general, then we represent F as the sum $F = \bigcup_{n=1}^{\infty} F_n$, where F_n is the intersection of closed sets F and $[-n, n]$. According to what has been proved in the first step every F_n is measurable (for it is closed and bounded) and therefore by Theorem 8.3 so is the set F. The proof of the theorem is complete.

Theorem 8.5. If a set E is measurable, then so is its complement CE.

Proof. By the definition of the measurability of E, for any n there is an open set G_n containing E for which

$$|G_n \setminus E|^* < \frac{1}{n}. \tag{8.10}$$

Let $F_n = CG_n$. Since $CE_1 \setminus CE_2 = E_2 \setminus E_1$ for any sets E_1 and E_2 (check this yourself), $CE \setminus CG_n = G_n \setminus E$ and therefore $CE \setminus F_n = G_n \setminus E$. It follows from the last equation that for any n

$$CE \setminus \bigcup_{k=1}^{\infty} F_k \subset G_n \setminus E. \tag{8.11}$$

(Recall that the notation $E_1 \subset E_2$ means that E_1 is in E_2.)

From (8.11) and from Property 1° of the outer measure we have for any n

$$|CE \setminus \bigcup_{k=1}^{\infty} F_k|^* \leqslant |G_n \setminus E|^*$$

and from the last inequality and from (8.10) we have

$$|CE \setminus \bigcup_{k=1}^{\infty} F_k|^* < \frac{1}{n}$$

(for any n). But this means that the outer measure and therefore the measure of the set $E_0 = CE \setminus \bigcup_{k=1}^{\infty} F_k$ equals zero, i.e. the set CE equals the sum of the measurable sets E_0 and $\bigcup_{k=1}^{\infty} F_l$ (the last set is measurable by virtue of Theorems 8.4 and 8.3). Thus the theorem is proved.

Corollary. *For a set E to be measurable, it is necessary and sufficient that given any positive number ε we should be able to find a closed set F contained in E such that the outer measure of the difference $E \setminus F$ is less than ε.*

Proof. The measurability of the set E is equivalent to the measurability of CE (Theorem 8.5), i.e. equivalent to the requirement that for any $\varepsilon > 0$ there should be an open set G containing CE and such that $|G \setminus CE|^* < \varepsilon$. But that requirement (by virtue of the identity $CE_1 \setminus CE_2 \equiv E_2 \setminus E_1$) is equivalent to the requirement that for any $\varepsilon > 0$ there should be a closed set $F = CG$ contained in E such that $|E \setminus F|^* = |CF \setminus CE|^* = |G \setminus CE|^* < \varepsilon$. Thus the corollary is proved.

Remark 1. The measurability condition contained in the corollary just proved may be accepted as another definition of measurability equivalent to the definition formulated at the beginning of this subsection.

Theorem 8.6. *The intersection of a finite or countable number of measurable sets is a measurable set.*

Proof. We shall denote the intersection of sets $E_1, E_2, \ldots$ by $\bigcap_{n=1}^{\infty} E_n$. By virtue of the identity $\bigcap_{n=1}^{\infty} E_n \equiv C[\bigcup_{n=1}^{\infty} CE_n]$ (check this identity yourself) the theorem being proved follows immediately from Theorem 8.3 and 8.5.

Theorem 8.7. *The difference of two measurable sets is a measurable set.*

A proof follows from the identity $A \setminus B \equiv A \cap (CB)$ and from Theorems 8.5 and 8.6.

We now proceed to prove the main theorem of measure theory.

Theorem 8.8. The measure of a sum of a finite or countable number of mutually disjoint measurable sets equals a sum of the measures of those sets.

Proof. Let $E = \bigcup\limits_{n=1}^{\infty} E_n$ such that the sets E_n being measurable and mutually disjoint. We consider separately *two cases.*

(1) First suppose that all E_n are *bounded.* Notice that for the case where all E_n are closed and there are a finite number of them the theorem being proved follows immediately from Property 3° of the outer measure (see Section 8.2.1).

Now let E_n be arbitrary bounded mutually disjoint sets.

By the corollary of Theorem 8.5, for any $\varepsilon > 0$ and for every n there is a closed set F_n contained in E_n such that* $|E_n \setminus F_n| \varepsilon/2^n$. Since all the sets F_n are bounded, closed and mutually disjoint, for any finite m, by virtue of the remark above

$$\left|\bigcup_{n=1}^{m} F_n\right| = \sum_{n=1}^{m} |F_n|. \tag{8.12}$$

On the other hand, from the equation $E_n = (E_n \setminus F_n) \cup F_n$ it follows (by virtue of Property 2° of the outer measure) that $|E_n| \leqslant \leqslant |E_n \setminus F_n| + |F_n| < |F_n| + \varepsilon/2^n$, so that

$$\sum_{n=1}^{m} |E_n| \leqslant \sum_{n=1}^{m} |F_n| + \varepsilon \tag{8.13}$$

(for any finite m). From (8.12) and (8.13) we conclude that for any finite m

$$\sum_{n=1}^{m} |E_n| \leqslant \left|\bigcup_{n=1}^{m} F_n\right| + \varepsilon. \tag{8.14}$$

Now we take into account the fact that the sum of all the sets F_n is contained in E. It follows that for any m

$$\left|\bigcup_{n=1}^{m} F_n\right| \leqslant |E|,$$

so that (by virtue of (8.14)) for any m

$$\sum_{n=1}^{m} |E_n| \leqslant |E| + \varepsilon. \tag{8.15}$$

Proceeding in (8.15) to the limit as $m \to \infty$, we get

$$\sum_{n=1}^{\infty} |E_n| \leqslant |E| + \varepsilon$$

* Since the measurability of all the sets occurring in the proof has already been established we may write everywhere simply measure instead of upper measure.

and therefore, in view of the arbitrariness of $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} |E_n| \leqslant |E|. \tag{8.16}$$

Now notice that from the equality of the sum $\bigcup\limits_{n=1}^{\infty} E_n$ to the set E and from Property 2° of the outer measure we obtain the inequality

$$|E| \leqslant \sum_{n=1}^{\infty} |E_n| \tag{8.17}$$

that has an opposite sense. From inequalities (8.16) and (8.17) we obtain the statement of the theorem being proved (for the case of bounded sets E_n).

(2) Now let the sets E_n *be not in general bounded.* Then we denote by E_n^k a bounded set $E_n^k = E_n \cap (k-1 \leqslant |x| < k)$ (recall that $\cap$ stands for "intersection").

From the equation $E = \bigcup\limits_{n=1}^{\infty} \bigcup\limits_{k=1}^{\infty} E_n^k$ and from the case considered above it follows that

$$|E| = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} |E_n^k| = \sum_{n=1}^{\infty} |E_n|.$$

This completes the proof of the theorem.

Remark 2. The fundamental property of measure established by Theorem 8.8 is called the *σ-additivity* of measure.

We introduce a new concept to state yet another property of measure.

Definition 2. A set E will be said to be a set of the type G_δ, *if E can be represented as an intersection of a countable number of open sets* G_n; *and* a set of the type F_σ, *if E can be represented as a sum of a countable number of closed sets* F_n.

Theorem 8.9. If a set E is measurable, then there are a set E_1 of the type F_σ contained in E and a set E_2 of the type G_δ containing E for which $|E_1| = |E| = |E_2|$.

Proof. By the measurability of E and the corollary of Theorem 8.5, for any n there are an open set G_n containing E and a closed set F_n contained in E such that

$$|E - F_n| < \frac{1}{n}, \quad |G_n \setminus E| < \frac{1}{n}. \tag{8.18}$$

Set $E_1 = \bigcup\limits_{n=1}^{\infty} F_n$, $E_2 = \bigcap\limits_{n=1}^{\infty} G_n$. Since for any n

$$E \setminus E_1 \subset E \setminus F_n, \quad E_2 \setminus E \subset G_n \setminus E,$$

by (8.18) and Property 1° of the outer measure

$$|E \setminus E_1| < \frac{1}{n}, \quad |E_2 \setminus E| < \frac{1}{n}.$$

By virtue of the arbitrariness of n it follows that $|E \setminus E_1| = 0$ and $|E_2 \setminus E| = 0$. Thus the theorem is proved.

Remark 3. Note that there are *nonmeasurable* sets. To construct them it suffices to take into consideration that on a unit circle there are a countable number of mutually disjoint and congruent* sets whose union is equal to the set of all the points of that circle. Such are the set E_0 of all the points of a circle any two of which cannot be made to coincide by rotation through an angle of $n \cdot \alpha$, where n is any integer and α is a fixed *irrational* number, and all the sets E_n obtained from E_0 by rotation through an angle of $n \cdot \alpha$. Were E_0 measurable, then so would be all the sets E_n, $|E_n| = |E_0|$ for all integers n. But then by virtue of Theorem 8.8 we would get $2\pi = \sum\limits_{n=-\infty}^{\infty} |E_n|$, which is impossible for any value of $|E_n|$.

8.3. MEASURABLE FUNCTIONS

8.3.1. The measurable function. Let us agree to mean by an *extended* number line the ordinary number line $-\infty < x < \infty$ complemented with two new elements, $-\infty$ and $+\infty$. To extend the arithmetic operations to include the extended number line, let us agree to assume that $a + (+\infty) = +\infty$, $a + (-\infty) = -\infty$ (for any finite a); $(+\infty) + (+\infty) = +\infty$, $-\infty + (-\infty) = -\infty$; $(+\infty) - a = +\infty$, $(-\infty) - a = -\infty$ (for any finite a), $(+\infty) - (-\infty) = +\infty$, $-\infty - (+\infty) = -\infty$; $a \cdot (+\infty) = +\infty$ for $a > 0$, $0 \cdot (+\infty) = 0$, $a \cdot (+\infty) = -\infty$ for $a < 0$; $(+\infty)(+\infty) = +\infty$, $(+\infty) \cdot (-\infty) = -\infty$, $(-\infty) \cdot (-\infty) = +\infty$, $0 \cdot (-\infty) = 0$, $a \cdot (-\infty) = -\infty$ for $a > 0$, $a \cdot (-\infty) = +\infty$ for $a < 0$; $\pm\infty/a = (\pm\infty) \cdot 1/a$ for any finite $a \neq 0$, $a/\pm\infty = 0$ for any finite a.

Only the following operations remain undefined:
$(+\infty) + (-\infty)$, $(+\infty) - (+\infty)$, $(-\infty) - (-\infty)$, $\pm\infty/\pm\infty$.

Throughout the remainder of this chapter we shall discuss functions defined on *measurable* sets of the *ordinary* number line and taking values on the *extended* number line.

$$f(x) = \begin{cases} -\infty & \text{when } x < -1, \\ 0 & \text{when } -1 \leqslant x \leqslant 1, \\ +\infty & \text{when } x > 1 \end{cases}$$

may serve as an example of such a function.

* In this case the term "congruent" must be taken to apply to two sets one of which may be made to coincide with the other by rotation through some angle in the plane of a circle.

Let us agree to denote henceforth by E [f satisfies condition A] a set of all values of x in E for which $f(x)$ satisfies condition A.

For example, $E[f \geqslant a]$ is a set of those values of x in E for which $f(x) \geqslant a$.

Definition. *A function $f(x)$ defined on a measurable set E is said to be* measurable *on that set if for any real number a the set $E[f \geqslant a]$ is measurable.*

Theorem 8.10. *For the function $f(x)$ to be measurable on a set E it is necessary and sufficient that one of the following three sets:*

$$E[f > a], \quad E[f < a], \quad E[f \leqslant a] \tag{8.19}$$

should be measurable for any real a.

Proof. (1) From the definition of measurability of the function $f(x)$, from the elementary relations

$$E[f > a] = \bigcup_{n=1}^{\infty} E\left[f \geqslant a + \frac{1}{n}\right],$$

$$E[f \geqslant a] = \bigcap_{n=1}^{\infty} E\left[f > a - \frac{1}{n}\right]$$

and from Theorems 8.3 and 8.6 it follows that the measurability (for any real a) of the set $E[f > a]$ is a necessary and sufficient condition for the function $f(x)$ to be measurable on the set E.

(2) From the relation $E[f < a] = E \setminus E[f \geqslant a]$ and from Theorems 8.3 and 8.7 it follows that the measurability (for any real a) of the set $E[f < a]$ is a necessary and sufficient condition for $f(x)$ to be measurable on E.

(3) Finally, from the relation $E[f \leqslant a] = E \setminus E[f > a]$, from Theorems 8.3 and 8.7, and from what has been proved in (1) it follows that the measurability (for any real a) of the set $E[f \leqslant a]$ is a necessary and sufficient condition for $f(x)$ to be measurable on E. Thus the theorem is proved.

Remark. By virtue of Theorem 8.10 the measurability (for any real a) of any of the three sets (8.19) may be accepted as another definition of the measurability of the function $f(x)$ on a set E equivalent to the definition formulated above.

8.3.2. Properties of measurable functions.

1°. *If a function $f(x)$ is measurable on a set E, then it is measurable on any measurable part E_1 of E.*

The proof follows immediately from the identity $E_1[f \geqslant a] = E_1 \cap E[f \geqslant a]$ and from Theorem 8.6.

2°. *If a set E is a finite or countable sum of measurable sets E_n and if $f(x)$ is measurable on every set E_n then it is measurable on the set E.*

The proof follows immediately from the identity $E[f \geqslant a] = \bigcup_{n=1}^{\infty} E_n[f \geqslant a]$ and from Theorem 8.3.

3°. *Any function $f(x)$ is measurable on a set E of measure zero.*

Indeed, any subset of a set of measure zero is measurable and has measure zero.

Definition 1. Two functions $f(x)$ and $g(x)$ defined on a measurable set E are said to be **equivalent** *on that set if the set $E[f \neq g]$ is of measure zero.*

The notation $f \approx g$ is often used to denote functions $f(x)$ and $g(x)$ that are equivalent (on a set E).

4°. *If $f(x)$ and $g(x)$ are equivalent on a set E and $f(x)$ is measurable on E, then so is $g(x)$.*

Proof. Put $E_0 = E[f \neq g]$, $E_1 = E \setminus E_0$. Since on E_1 $g(x)$ coincides with $f(x)$, $g(x)$ is measurable on E_1 (by virtue of Property 1°). According to Property 3° $g(x)$ is measurable on E_0 as well as, therefore, according to Property 2° $g(x)$ is also measurable on E.

Definition 2. We shall say that some property A is true **almost everywhere** *on a set E if the set of points of E on which that property is false is of measure zero.*

Corollary of Property 4°. If a function $f(x)$ is continuous almost everywhere on a measurable set E, then $f(x)$ is measurable on E.

Proof. First notice that if $f(x)$ is continuous on a closed set F then $f(x)$ is measurable on F, for $F[f \geqslant a]$ is closed for any real a and, therefore, measurable. Suppose that $f(x)$ is continuous almost everywhere on an arbitrary measurable set E and denote by R a subset of all discontinuity points of $f(x)$ of measure zero.

By virtue of Properties 2° and 3° it suffices to prove the measurability of $f(x)$ on the set $E_1 = E \setminus R$. By Theorem 8.9 there is a set E_2 of the type F_σ (see Section 8.2.2) contained in E_1 such that $|E_2| = |E_1| = |E|$. By virtue of the same Properties 2° and 3° it suffices to prove that $f(x)$ is measurable on E_2. But E_2 (as a set of the type F_σ) can be represented as a countable sum of closed sets F_n on each of which $f(x)$ is continuous and therefore (in view of the above remark) measurable. But then by Property 2° $f(x)$ is measurable on E_2.

Remark. We stress that the continuity of a function $f(x)$ almost everywhere on a set E should be distinguished from the equivalence of $f(x)$ on E to a continuous function. Thus the Dirichlet function $f(x) = 1$, if x is rational; $f(x) = 0$, if x is irrational and continuous at no point of $[0, 1]$ (see Chapter 4 of [1]); but it is equivalent on $[0, 1]$ to the continuous function $g(x) \equiv 0$ since $f(x) \neq g(x)$ only on the set of all rational points of $[0, 1]$ which is countable and is therefore of measure zero*.

8.3.3. Arithmetic operations on measurable functions. We first prove the following lemma.

* The fact that a countable set of points is of measure zero follows from Theorem 8.8 and from the fact that the measure of a set consisting of a single point is zero.

Lemma 1. (1) *If a function $f(x)$ is measurable on a set E, then so is the function $|f(x)|$. (2) If $f(x)$ is measurable on E, and C is any constant, then either of the functions $f(x) + C$ and $C \cdot f(x)$ is measurable on E. (3) If $f(x)$ and $g(x)$ are measurable on E, then the set $E[f > g]$ is measurable.*

Proof. (1) It suffices to consider that for any nonnegative a

$$E[|f| \geqslant a] = E[f \geqslant a] \cup E[f \leqslant -a]$$

and use Theorem 8.3. If, however, $a < 0$, then $E[|f| > a]$ coincides with E and is also measurable.

(2) It suffices to use for any real a the relations

$$E[f + C \geqslant a] = E[f \geqslant a = C],$$

$$E[C \cdot f \geqslant a] = \begin{cases} E\left[f \geqslant \frac{a}{C}\right] & \text{when } C > 0, \\ E\left[f \leqslant \frac{a}{C}\right] & \text{when } C > 0. \end{cases}$$

If, however, $C = 0$, then $C \cdot f(x) \equiv 0$ and is also measurable.

(3) Let $\{r_k\}$ be all rational points of an infinite straight line $(-\infty, \infty)$. It suffices to consider that

$$E[f > g] = \bigcup_{k=1}^{\infty} (E[f > r_k] \cap E[g < r_k])$$

and use Theorems 8.3 and 8.6. Thus the lemma is proved.

On the basis of Lemma 1 we prove the following theorem.

Theorem 8.11. If functions $f(x)$ and $g(x)$ take finite values on a set E and are measurable on it, then each of the functions $f(x) - g(x)$, $f(x) + g(x)$, $f(x) \cdot g(x)$, and $f(x)/g(x)$ (for the quotient $f(x)/g(x)$ we require in addition that all the values of $g(x)$ should be nonzero) is measurable on E.

Proof. (1) To prove that the difference $f(x) - g(x)$ is measurable it suffices to notice that for any real a the set $E[f - g > a]$ coincides with the measurable (by Lemma 1) set $E[f > g + a]$.

(2) To prove that the sum $f(x) + g(x)$ is measurable it suffices to consider that $f + g = f - (-g)$ and that the function $-g(x)$ is measurable by Lemma 1.

(3) To prove that a product of two measurable functions is measurable we first show that the square of a measurable function is a measurable function. Indeed, if $a < 0$, then the set $E[f^2 > a]$ coincides with E and is therefore measurable. If, however, $a \geqslant 0$, then the set $E[f^2 > a]$ coincides with the measurable (by Lemma 1) set $E[|f| > \sqrt{a}]$. From the measurability of the square of a measurable function and from the measurability of the sum and the difference of measurable functions, by virtue of the relation $f \cdot g =$

$= 1/4\,(f+g)^2 - 1/4\,(f-g)^2$ it follows that the product $f(x)\,g(x)$ is measurable.

(4) By virtue of the measurability of a product of two measurable functions, to prove that the quotient f/g is measurable it suffices to establish that $1/g$ is measurable, but this follows from Theorems 8.3 and 8.6 and from the relation

$$E\left[\frac{1}{g}>a\right]=\begin{cases} E\,[g>0]\cap E\left[g<\frac{1}{a}\right] & \text{when } a>0,\\ E\,[g>0] & \text{when } a=0,\\ E\,[g>0]\cup E\left[g<\frac{1}{a}\right] & \text{when } a<0.\end{cases}$$

The proof of the theorem is complete.

8.3.4. Sequences of measurable functions. We prove several important statements relating to sequences of measurable functions.

Theorem 8.12. If $\{f_n(x)\}$ is a sequence of functions measurable on a set E, then both the lower and upper limits of the sequence are functions measurable on E.*

Proof. We first show that if a sequence $\{g_n(x)\}$ consists of functions measurable on E, then either of the functions** $\varphi(x) = \inf\limits_n g_n(x)$ and $\psi(x) = \sup\limits_n g_n(x)$ is measurable on E. It suffices to take into consideration the relations

$$E\,[\varphi<a]=\bigcup_{n=1}^{\infty} E\,[g_n<a],$$

$$E\,[\psi>a]=\bigcup_{n=1}^{\infty} E\,[g_n>a]$$

and use Theorem 8.3.

Now denote the lower and upper limits of $\{f_n(x)\}$ respectively by $\underline{f}(x)$ and $\overline{f}(x)$. To prove that $\underline{f}(x)$ and $\overline{f}(x)$ are measurable on E it suffices to notice that

$$\underline{f}(x)=\sup_{n\geqslant 1}\{\inf_{k\geqslant n} f_k(x)\},\quad \overline{f}(x)=\inf_{n\geqslant 1}\{\sup_{k\geqslant n} f_k(x)\}$$

and use the statement proved above. Thus the theorem is proved.

---

* In Chapter 3 of [1] we have proved the existence of the lower and upper limits in any bounded sequence. Here we agree to assume that if a sequence is not bounded below (upper), then its lower (upper) limit is equal to $-\infty$ $(+\infty)$.

** The notation $\varphi(x) = \inf\limits_n g_n(x)$ means that at every point x the value of $\varphi(x)$ is the infimum of the values $g_1(x), g_2(x), \ldots$ at that point. The notation $\psi(x) = \sup\limits_n g_n(x)$ has a similar meaning.

Theorem 8.13. If a sequence $\{f_n(x)\}$ of functions measurable on a set E converges almost everywhere on E to a function $f(x)$, then $f(x)$ is measurable on E.

Proof. In the case where a function $\{f_n(x)\}$ converges to $f(x)$ *everywhere* rather than almost everywhere on E the statement of the theorem on the measurability of $f(x)$ follows at once from Theorem 8.12. If, however, $\{f_n(x)\}$ converges to $f(x)$ everywhere on E, except the set E_0 of measure zero, then $f(x)$ is measurable on $E \setminus E_0$ by virtue of Theorem 8.12 and is measurable on E_0 as a set of measure zero (Property 3° of Section 8.3.2) and hence on the set $E = (E \setminus E_0) \cup E_0$ (by virtue of Property 2° of Section 8.3.2). Thus the theorem is proved.

Now we introduce an important concept of convergence of a sequence *in measure* on a given set.

Definition. Let functions $f_n(x)$ $(n = 1, 2, \ldots)$ and $f(x)$ be measurable on a set E and take on finite values almost everywhere on E. The sequence $\{f_n(x)\}$ is said to converge to $f(x)$ in measure *on E if for any positive number ε*

$$\lim_{n\to\infty} |E[|f-f_n| \geqslant \varepsilon]| = 0, \tag{8.20}$$

i.e. if for any positive ε and δ there is N such that when $n \geqslant N$ we have $|E[|f - f_n| \geqslant \varepsilon]| < \delta$.

H. Lebesgue proved the following theorem.

Theorem 8.14. Let E be a measurable set of finite measure and let functions $f_n(x)$ $(n = 1, 2, \ldots)$ and $f(x)$ be measurable on E and take on finite values almost everywhere on E. Then the convergence of $\{f_n(x)\}$ to $f(x)$ almost everywhere on E implies the convergence of $\{f_n(x)\}$ to $f(x)$ also in measure on E.

Proof. Set $A = E[|f| = +\infty]$, $A_n = E[|f_n| = +\infty]$, $B = E \setminus E[\lim\limits_{n\to\infty} f_n = f]$, $C = A + B + \bigcup\limits_{n=1}^{\infty} A_n$. Then under the condition of the theorem $|C| = 0$, the sequence $\{f_n(x)\}$ converges to $f(x)$ everywhere outside the set C and all the functions $f_n(x)$ and $f(x)$ have finite values.

For an arbitrary $\varepsilon > 0$, set $E_n = E[|f - f_n| \geqslant \varepsilon]$, $R_n = \bigcup\limits_{k=n}^{\infty} E_k$. Then, since E_n is contained in R_n, we have $|E_n| \leqslant |R_n|$, and to prove (8.20) *it suffices to establish that* $|R_n| \to 0$ *as* $n \to \infty$.

Denote by R the intersection of all sets $R_1, R_2, \ldots$ to show that $|R_n| \to |R|$ as $n \to \infty$. By construction R_{n+1} is contained in R_n for every n and therefore for every n

$$R_n \setminus R = \bigcup_{k=n}^{\infty} (R_k \setminus R_{k+1}),$$

with the sets under the summation sign mutually disjoint. But then, by Theorem 8.8, for every n

$$|R_n \setminus R| = \sum_{k=n}^{\infty} |R_k \setminus R_{k+1}| \tag{8.21}$$

and, by the convergence of the series

$$|R_1 \setminus R| = \sum_{k=1}^{\infty} |R_k \setminus R_{k+1}|,$$

the remainder of the series (8.21) converges to zero as $n \to \infty$. So $|R_n \setminus R| \to 0$ as $n \to \infty$. But by virtue of the relation $|R_n| = |R_n \setminus R| + |R|$ this means that $|R_n| \to |R|$ as $n \to \infty$.

Now, to prove (8.20) we establish that $|R| = 0$. To do this it is in turn *sufficient to establish that R is contained in C.*

Let x_0 be any point outside C. Then for the arbitrary $\varepsilon > 0$ we have fixed there is $N(x_0, \varepsilon)$ such that $|f_n(x_0) - f(x_0)| < \varepsilon$ when $n \geqslant N(x_0, \varepsilon)$. But this means that when $n \geqslant N(x_0, \varepsilon)$ the point x_0 is not in E_n and, the more so, it is not in R_n or in the set R which is the intersection of all R_n.

So any point x_0 which is outside C is also outside R. But this means that R is contained in C. Thus the theorem is proved.

Remark. We stress that convergence of a sequence $\{f_n(x)\}$ to $f(x)$ on E in measure implies neither the convergence of $\{f_n(x)\}$ to $f(x)$ almost everywhere on E, nor even the convergence of $\{f_n(x)\}$ to $f(x)$ *at at least one point of E.* It suffices to consider the example constructed in Section 1.2.3. The sequence $\{f_n(x)\}$ in that example diverges at each point of the closed interval $[0, 1]$, but since every function $f_n(x)$ is nonzero only on the closed interval I_n whose length tends to zero as $n \to \infty$ the sequence $\{f_n(x)\}$ converges to the function $f(x) \equiv 0$ in measure on $[0, 1]$.

Nevertheless Frigyes Riesz* proved the following theorem.

Theorem 8.15. Let E be a measurable set of finite measure and let functions $f_n(x)$ $(n = 1, 2, \ldots)$ and $f(x)$ be measurable on E and take on finite values almost everywhere on E. Then if $\{f_n(x)\}$ converges to $f(x)$ in measure on E, we can choose a subsequence of $\{f_n(x)\}$ converging to $f(x)$ almost everywhere on E.

Proof. We may assume without loss of generality that $f_n(x)$ and $f(x)$ take on finite values everywhere rather than almost everywhere on E (otherwise the same sets A and A_n would be introduced that were in proving the previous theorem and all the reasoning would be carried out for the set $E \setminus A \bigcup_{n=1}^{\infty} A_n$). It follows from the convergence of $\{f_n(x)\}$ to $f(x)$ in measure on E that, given any k, we can find n_k such that for the measure of the set $E_k = E\,[|f -$

* Frigyes Riezs (1880-1956), a Hungarian mathematician.

$-f_{n_k}|\geqslant 1/k\}$ we have $|E_k|\leqslant 1/2^k$. Set, as in the proof of the previous theorem, $R_n=\bigcup_{k=n}^{\infty} E_k$, $R=\bigcap_{n=1}^{\infty} R_n$. Then by the property of the outer measure (see Section 8.2.1) $|R_n|=\sum_{k=n}^{\infty}|E_k|$, so that $|R_n|\leqslant$ $\leqslant \sum_{k=n}^{\infty} 1/2^k=1/2^{n-1}$. Thus $|R_n|\to 0$ as $n\to\infty$. As in the previous theorem, we can prove that $|R_n|\to|R|$ as $n\to\infty$. This yields $|R|=0$.

It remains to prove that everywhere outside R the sequence $\{f_{n_k}(x)\}$ converges to $f(x)$. Let x be an arbitrary point of $E\setminus R$. Then x is not in the set R_N for some $N=N(x)$. But this means that x is not in E_k when $k\geqslant N(x)$. In other words, $|f(x)-f_{n_k}(x)|<$ $<1/k$ when $k\geqslant N(x)$. Thus the theorem is proved.

8.4. THE LEBESGUE INTEGRAL

8.4.1. The Lebesgue integral of a bounded function. A *subdivision* of a measurable set E is any family T of a finite number of measurable and mutually disjoint sets $E_1, E_2, \ldots, E_n$ of E adding up to the set E.

To denote a subdivision of a set E we shall use the symbol $T=$ $=\{E_k\}_{k=1}^{n}$ or $T=\{E_k\}$ for short.

Consider on a measurable set E of finite measure an arbitrary bounded function $f(x)$. For an arbitrary subdivision $T=\{E_k\}$ of E we denote by M_k and m_k respectively the supremum and infimum of $f(x)$ on a subset E_k and define the two sums

$$S_T=\sum_{k=1}^{n} M_k|E_k| \quad\text{and}\quad s_T=\sum_{k=1}^{n} m_k|E_k|$$

called respectively the *upper* and *lower* sums of the subdivision $T=\{E_k\}$.

Note at once that for any subdivision $T=\{E_k\}$

$$s_T\leqslant S_T. \tag{8.22}$$

For any function $f(x)$ bounded on a set of finite measure E both the set of all upper sums $\{S_T\}$ and the set of all lower sums $\{s_T\}$ (corresponding to all possible subdivisions $T=\{E_k\}$ of the set E) are bounded. There is therefore an infimum of the set $\{S_T\}$, which we designate $\overline{I}$ and call the *upper Lebesgue integral*, and a supremum of $\{s_T\}$, which we designate $\underline{I}$ and call the *lower Lebesgue integral*.

Definition. *A function $f(x)$ bounded on a set of finite measure E is said to be* (Lebesgue) integrable *on that set if $\underline{I}=\overline{I}$, i.e. if the upper and lower Lebesgue integrals of the function coincide.*

The number $\underline{I} = \overline{I}$ is called the Lebesgue integral of $f(x)$ over E and designated

$$\int_E f(x)\,dx.$$

We discuss some properties of the upper and lower sums and of the upper and lower Lebesgue integrals.

Let us say that the subdivision $T^* = \{E_i^*\}_{i=1}^m$ is a *refinement* of the subdivision $T = \{E_k\}_{k=1}^n$ if for any i $(i = 1, 2, \ldots, m)$ there is $\nu(i)$ satisfying the inequalities $1 \leqslant \nu(i) \leqslant n$ such that E_i^* is contained in $E_{\nu(i)}$.

The integer $\nu(i)$ may turn out to be the same for different i, the sum of the sets E_i^* over all i for which $\nu(i)$ equals the same integer k is obviously equal to a set E_k, i.e.

$$\bigcup_{\nu(i)=k} E_i^* = E_k \tag{8.23}$$

Let us also say that the subdivision $\hat{T} = \{E_i\}$ is a *product* of subdivisions $T_1 = \{E_p^{(1)}\}$ and $T_2 = \{E_q^{(2)}\}$ if $\hat{T}$ consists of the sets E_i which are the intersections of all possible pairs of sets $E_p^{(1)}$ and $E_q^{(2)}$, i.e. if every E_i is equal to $E_p^{(1)} \cap E_q^{(2)}$, all possible combinations of p and q exhausted.

Obviously the product $\hat{T}$ of two subdivisions T_1 and T_2 is a refinement of either of the subdivisions T_1 and T_2 (any other subdivision of T which is a refinement of both T_1 and T_2 is itself a refinement of $\hat{T}$).

The following properties of the upper and lower sums and of the upper and lower integrals are valid.

1°. *If a subdivision T^* is a refinement of a subdivision T, then* $s_T \leqslant s_{T^*}$, $S_{T^*} \leqslant S_T$.

Proof. We give a proof for the *upper* sums (since for the lower sums the proof is quite similar). Let $T^* = \{E_i^*\}_{i=1}^m$ be a refinement $T = \{E_k\}_{k=1}^n$ and let M_i^* be the supremum of $f(x)$ on a set E_i^* $(i = 1, 2, \ldots, m)$ and M_k the supremum of $f(x)$ on a set E_k $(k = 1, 2, \ldots, n)$.

By the definition of refinement for every number i $(i = 1, 2, \ldots, m)$ there is a corresponding integer $\nu(i)$ satisfying $1 \leqslant \nu(i) \leqslant n$ such that E_i^* is contained in $E_{\nu(i)}$, the sum of the sets E_i^* over all i for which $\nu(i)$ equals the same k satisfying equation (8.23). It should be added that for all i for which $\nu(i)$ equals the same integer k

$$M_i^* \leqslant M_k \tag{8.24}$$

(for the supremum on a subset does not exceed the supremum on the whole of the set).

From the definition of the upper sum and from relations (8.23) and (8.24) we get*

$$S_{T^*} = \sum_{i=1}^{m} M_i^* |E_i^*| = \sum_{k=1}^{n} [\sum_{\nu(i)=k} M_i^* |E_i^*|] \leqslant$$

$$\leqslant \sum_{k=1}^{n} M_k [\sum_{\nu(i)=k} |E_i^*|] = \sum_{k=1}^{n} M_k |E_k| = S_T.$$

2°. *For two quite arbitrary subdivisions T_1 and T_2, $s_{T_1} \leqslant S_{T_2}$.*

Proof. Let $\tilde{T}$ be the product of subdivisions T_1 and T_2. Since $\tilde{T}$ is a refinement of either of the subdivisions T_1 and T_2, by Property 1°

$$s_{T_1} \leqslant s_{\tilde{T}}, \quad S_{\tilde{T}} \leqslant S_{T_2}. \tag{8.25}$$

From inequalities (8.25) and (8.22) it follows that $s_{T_1} \leqslant S_{T_2}$.

3°. *The upper and lower Lebesgue integrals are connected by the relation $\underline{I} \leqslant \overline{I}$.*

Proof. Fix an *arbitrary* subdivision T_2. Since for any subdivision T_1 (by Property 2°) $s_{T_1} \leqslant S_{T_2}$, the number S_{T_2} is *one* of the upper bounds of the set $\{s_{T_1}\}$ of all the lower sums, and therefore the *supremum* $\underline{I}$ of that set satisfies the inequality $\underline{I} \leqslant S_{T_2}$. Since the last inequality is true for an arbitrary subdivision T_2, the number $\underline{I}$ is *one* of the lower bounds of the set $\{S_{T_2}\}$ of all the upper sums, and therefore the infimum $\overline{I}$ of that set satisfies the condition $\underline{I} \leqslant \overline{I}$.

Corollary. Any Riemann integrable function is Lebesgue integrable, the Lebesgue and Riemann integrals of such a function coinciding.

Proof. Let $f(x)$ be Riemann integrable on $E = [a, b]$ (and therefore bounded on that interval). For such a function we denote by $\underline{I}$ and $\overline{I}$ the lower and upper Lebesgue integrals and by $\underline{I}_R$ and $\overline{I}_R$ the lower and upper Darboux integrals (see Chapter 10 of [1]) to obtain the following inequalities**

$$\underline{I}_R \leqslant \underline{I} \leqslant \overline{I} \leqslant \overline{I}_R. \tag{8.26}$$

If a function is Riemann integrable, then for it $\underline{I}_R = \overline{I}_R$ and by (8.26) therefore $\underline{I} = \overline{I}$, i.e. that function is Lebesgue integrable. Moreover, when $\underline{I}_R = \overline{I}_R$ (8.26) implies $\underline{I}_R = \underline{I} = \overline{I} = \overline{I}_R$, i.e. they imply the coincidence of the Riemann and Lebesgue integrals,

---

* We take into account that from (8.23) and from the sets E_i^* being mutually disjoint it follows, by Theorem 8.8, that $\sum_{\nu(i)=k} |E_i^*| = |E_k|$.

** Because any subdivision $E = [a, b]$ into subintervals is included in the class of subdivisions of the set E in the sense of Lebesgue.

for the former is equal to the number $\underline{I}_R = \overline{I}_R$ and the latter to $\underline{I} = \overline{I}$.

In the next subsection we shall show that the class of Lebesgue integrable functions is wider than the class of Riemann integrable functions. This will clarify the appropriateness of introducing measurable functions.

8.4.2. The class of Lebesgue integrable bounded functions. We prove the following *main* theorem.

Theorem 8.16. Whatever a measurable set E of finite measure may be, any function f (x) bounded and measurable on E is integrable on it.

Proof. Construct a special subdivision of a set E called *Lebesgue* subdivision. We denote by M and m the supremum and infimum of $f(x)$ on E and divide the interval $[m, M]$, using the points $m = y_0 < y_1 < y_2 < \ldots < y_n = M$, into subintervals $[y_{k-1}, y_k]$ $(k = 1, 2, \ldots, n)$ and denote by δ the length of the largest of the subintervals, i.e. set

$$\delta = \max_{k=1, 2, \ldots, n} (y_k - y_{k-1}).$$

A *Lebesgue* subdivision of a set E is a subdivision $T = \{E_k\}_{k=1}^{n}$ in which $E_1 = E\,[y_0 \leqslant f \leqslant y_1]$, $E_k = E\,[y_{k-1} < f \leqslant y_k]$ for $k = 2, 3, \ldots, n$.

Let S_T and s_T be the upper and lower sums corresponding to a Lebesgue subdivision T and called the *upper* and *lower Lebesgue sums*. Notice that for any k $(k = 1, 2, \ldots, n)$

$$y_{k-1} \leqslant m_k \leqslant M_k \leqslant y_k \tag{8.27}$$

where M_k and m_k denote the supremum and infimum of $f(x)$ on a subset E_k. Multiplying inequalities (8.27) by the measure $[E_k]$ of the set E_k and then summing them over all $k = 1, 2, \ldots, n$ we have

$$\sum_{k=1}^{n} y_{k-1} |E_k| \leqslant s_T \leqslant S_T \leqslant \sum_{k=1}^{n} y_k |E_k|.$$

From these inequalities we conclude that

$$0 \leqslant S_T - s_T \leqslant \sum_{k=1}^{n} y_k |E_k| - \sum_{k=1}^{n} y_{k-1} |E_k| =$$

$$= \sum_{k=1}^{n} (y_k - y_{k-1}) |E_k| < \delta |E|. \tag{8.28}$$

Since for any subdivision T we have $s_T \leqslant \underline{I} \leqslant \overline{I} \leqslant S_T$, from (8.28) we get

$$0 \leqslant \overline{I} - \underline{I} < \delta |E|. \tag{8.29}$$

Since $\delta > 0$ may be fixed as small as we please, it follows from (8.29) that $\underline{I} = \overline{I}$. Thus the theorem is proved.

Remark 1. In Supplement 2 to this chapter we prove that the measurability of a function $f(x)$ bounded on a measurable set E is not only a sufficient but also necessary condition for that function to be Lebesgue integrable on E.

Remark 2. Let ξ_k ($k = 1, 2, \ldots, n$) be an arbitrary element of a subset E_k of a Lebesgue subdivision T. A sum $\sigma_T(\xi_k, f) = \sum_{k=1}^{n} f(\xi_k) \cdot |E_k|$ will be called a *Lebesgue integral sum* of $f(x)$. Since under an arbitrary choice of points ξ_k on the sets E_k that sum is contained between the lower and upper sums of the corresponding Lebesgue subdivision T, it follows from inequality (8.28) that $\sigma_T(\xi_k, f)$ (together with S_T and s_T) tends, as $\delta \to 0$, to the Lebesgue integral $\underline{I} = \overline{I} = \int\limits_E f(x)\,dx$.

8.4.3. Properties of the Lebesgue integral of a bounded function.

1°. $\int\limits_E 1\,dx = |E|$.

To prove this it suffices to notice that for the function $f(x) \equiv 1$ both the upper and the lower sum of any subdivision T of a set E is equal to $|E|$.

2°. *If a function $f(x)$ is bounded and integrable on a set E of finite measure and α is any real number, then the function $[\alpha \cdot f(x)]$ is also integrable on E, with*

$$\int\limits_E [\alpha \cdot f(x)]\,dx = \alpha \cdot \int\limits_E f(x)\,dx. \tag{8.30}$$

Proof. For an arbitrary subdivision $T = \{E_k\}$ of a set E, denote the upper and lower sums of the function $f(x)$ by S_T and s_T and the upper and lower sums of the function $[\alpha \cdot f(x)]$ by $S_T^{(\alpha)}$ and $s_T^{(\alpha)}$. Then clearly

$$S_T^{(\alpha)} = \begin{cases} \alpha S_T & \text{when } \alpha \geqslant 0, \\ \alpha s_T & \text{when } \alpha < 0, \end{cases} \quad s_T^{(\alpha)} = \begin{cases} \alpha \cdot s_T & \text{when } \alpha \geqslant 0. \\ \alpha \cdot S_T & \text{when } \alpha < 0. \end{cases} \tag{8.31}$$

If we denote by $\overline{I}$ and $\underline{I}$ the upper and lower integrals of $f(x)$ and by $\overline{I}^{(\alpha)}$ and $\underline{I}^{(\alpha)}$ the upper and lower integrals of $[\alpha \cdot f(x)]$, then it follows from (8.31) that

$$\overline{I}^{(\alpha)} = \begin{cases} \alpha \cdot \overline{I} & \text{when } \alpha \geqslant 0, \\ \alpha \cdot \underline{I} & \text{when } \alpha < 0. \end{cases} \quad \underline{I}^{(\alpha)} = \begin{cases} \alpha \cdot \underline{I} & \text{when } \alpha \geqslant 0, \\ \alpha \cdot \overline{I} & \text{when } \alpha < 0. \end{cases} \tag{8.32}$$

By virtue of the integrability of $f(x)$

$$\underline{I} = \bar{I} = \int_E f(x)\, dx$$

and therefore from inequalities (8.32) it follows that for any α

$$\bar{I}^{(\alpha)} = \underline{I}^{(\alpha)} = \alpha \cdot \int_E f(x)\, dx.$$

This just means that the integral on the left of (8.30) exists and that equation (8.30) is true.

3°. *If either of the functions $f_1(x)$ and $f_2(x)$ is bounded and integrable on a set of finite measure E, then the sum of the functions $[f_1(x) + f_2(x)]$ is integrable on E, with*

$$\int_E [f_1(x) + f_2(x)]\, dx = \int_E f_1(x)\, dx + \int_E f_2(x)\, dx. \tag{8.33}$$

Proof. Set $f(x) = f_1(x) + f_2(x)$ and let $T = \{E_k\}$ be an arbitrary subdivision of E. For the function $f(x)$ denote the supremum and infimum on a subset E_k by M_k and m_k, the upper and lower sums of T by S_T and s_T, and the upper and lower Lebesgue integrals by $\bar{I}$ and $\underline{I}$. Similar quantities for the functions $f_1(x)$ and $f_2(x)$ we denote by the same symbols as for $f(x)$ but with the superscripts (1) and (2) respectively.

Notice that the *supremum (infimum) of a sum is not greater (not less) than the sum of the suprema (infima) of the summands.* It follows that for any k

$$m_k^{(1)} + m_k^{(2)} \leqslant m_k \leqslant M_k \leqslant M_k^{(1)} + M_k^{(2)}$$

and therefore for any subdivision T

$$s_T^{(1)} + s_T^{(2)} \leqslant s_T \leqslant S_T \leqslant S_T^{(1)} + S_T^{(2)}.$$

From these last inequalities it follows then that

$$\underline{I}^{(1)} + \underline{I}^{(2)} \leqslant \underline{I} \leqslant \bar{I} \leqslant \bar{I}^{(1)} + \bar{I}^{(2)}. \tag{8.34}$$

Since (due to the integrability of $f_1(x)$ and $f_2(x)$)

$$\underline{I}^{(1)} = \bar{I}^{(1)} = \int_E f_1(x)\, dx, \quad \underline{I}^{(2)} = \bar{I}^{(2)} = \int_E f_2(x)\, dx,$$

from (8.34) we get

$$\underline{I} = \bar{I} = \int_E f_1(x)\, dx + \int_E f_2(x)\, dx.$$

But this just means that the integral on the left of (8.33) exists and that equation (8.33) is true.

Corollary. Properties 2° and 3° lead directly to the linear property of the integral: *if either of the functions $f_1(x)$ and $f_2(x)$ is bounded and integrable on a set of finite measure E and if α and β are arbitrary real numbers, then the function $[\alpha \cdot f_1(x) + \beta \cdot f_2(x)]$ is integrable on E, with*

$$\int_E [\alpha f_1(x) + \beta f_2(x)]\, dx = \alpha \cdot \int_E f_1(x)\, dx + \beta \int_E f_2(x)\, dx.$$

4°. *If a function $f(x)$ is bounded and integrable on either of the disjoint sets of finite measure E_1 and E_2, then $f(x)$ is integrable on the sum E of the sets E_1 and E_2 as well, with*

$$\int_E f(x)\, dx = \int_{E_1} f(x)\, dx + \int_{E_2} f(x)\, dx. \tag{8.35}$$

This property is usually called the *additivity* of the integral.

Proof. Notice that the union of an arbitrary subdivision T_1 of the set E_1 and an arbitrary subdivision T_2 of E_2 forms a subdivision T of the set $E = E_1 \cup E_2$. Denote the upper sums of $f(x)$ corresponding to T_1, T_2 and T by S_{T_1}, S_{T_2} and S_T respectively, and the lower sums of $f(x)$ corresponding to T_1, T_2 and T by s_{T_1}, s_{T_2} and s_T respectively. Then clearly

$$S_T = S_{T_1} + S_{T_2}, \quad s_T = s_{T_1} + s_{T_2}. \tag{8.36}$$

Denote the upper and lower integrals of $f(x)$ over E_1 by $\bar{I}^{(1)}$ and $\underline{I}^{(1)}$, those over E_2 by $\bar{I}^{(2)}$ and $\underline{I}^{(2)}$ and those over E by $\bar{I}$ and $\underline{I}$.

From equations (8.36) and from the fact that the supremum (infimum) of a sum is not greater (not less) than the sum of the suprema (infima) of its terms we conclude that

$$\underline{I}^{(1)} + \underline{I}^{(2)} \leqslant \underline{I} \leqslant \bar{I} \leqslant \bar{I}^{(1)} + \bar{I}^{(2)}. \tag{8.37}$$

Since (due to the integrability of $f(x)$ on E_1 and on E_2) $\underline{I}^{(1)} = \bar{I}^{(1)} = \int_{E_1} f(x)\, dx$, $\underline{I}^{(2)} = \bar{I}^{(2)} = \int_{E_2} f(x)\, dx$, from (8.37) we get

$$\underline{I} = \bar{I} = \int_{E_1} f(x)\, dx + \int_{E_2} f(x)\, dx.$$

But this proves that the integral on the left of (8.35) exists and that equation (8.35) is true.

5°. *If either of the functions $f_1(x)$ and $f_2(x)$ is bounded and integrable on a set of finite measure E and if everywhere on that set $f_1(x) \geqslant$*

$\geqslant f_2(x)$, *then*

$$\int_E f_1(x)\,dx \geqslant \int_E f_2(x)\,dx. \tag{8.38}$$

Proof. Since all the lower sums of the function $F(x) = f_1(x) - f_2(x)$ are nonnegative, $\underline{I} \geqslant 0$. It follows that $\int_E F(x)\,dx = \int_E f_1(x)\,dx - \int_E f_2(x)\,dx \geqslant 0$ (the existence of the integral and the above equality follow from the linear property already proved). This proves (8.38).

8.4.4. The Lebesgue integral of a nonnegative unbounded function and its properties. We now proceed to define the Lebesgue integral for the case where the measurable function $f(x)$ is not bounded. We first *assume $f(x) \geqslant 0$ everywhere on a set of finite measure E.*

For any $N > 0$, we set

$$(f)_N(x) = \min\{N, f(x)\}, \tag{8.39}$$

$$I_N(f) = \int_E (f)_N(x)\,dx. \tag{8.40}$$

Notice that for any function $f(x)$ measurable on E the function (8.39) is also measurable* and therefore the integral (8.40) exists. Note also that from (8.39) and (8.40) it follows that $I_N(f)$ *increases as N increases.*

Definition. If there is a finite limit $I_N(f)$ as $N \to \infty$, then $f(x)$ is said to be (Lebesgue) **summable** *on a set E and the limit is said to be the* **integral** *of $f(x)$ over E and designated* $\int_E f(x)\,dx$. So by definition

$$\int_E f(x)\,dx = \lim_{N\to\infty} I_N(f).$$

We show that *if a function $f(x)$ nonnegative on a set E is summable on that set, then $f(x)$ may tend to $+\infty$ only on a subset of E of measure zero.* Indeed, set $E_0 = E[f = +\infty]$ and consider that from (8.40) and (8.39) (by Properties 4° and 5° of Section 8.4.3) we have a chain of inequalities

$$I_N(f) = \int_E (f)_N(x)\,dx \geqslant \int_{E_0} (f)_N(x)\,dx \geqslant \int_{E_0} N\,dx \geqslant N|E_0|.$$

---

* Since for any real a the following set is measurable:

$$E[(f)_N > a] = \begin{cases} E[f > a] & \text{when } a < N, \\ \text{an empty set} & \text{when } a \geqslant N. \end{cases}$$

But it follows from $I_N(f) \geqslant N \mid E_0 \mid$ that the assumption $\mid E_0 \mid > 0$ would result in the limit $\lim\limits_{N\to\infty} I_N(f)$ being equal to $+\infty$.

It should be added that *any function $f(x)$ is summable on a set of measure zero.* (This fact is obvious.)

Turning to the general properties of summable functions, first note that *Properties 2° to 5° established in Section 8.4.3 for bounded integrable functions hold for nonnegative summable functions.**

By way of illustration, we prove Property 3°. From (8.39) we have at once the following inequalities:

$$(f_1)_{N/2}(x) + (f_2)_{N/2}(x) \leqslant (f_1+f_2)_N(x) \leqslant (f_1)_N(x) + (f_2)_N(x)$$

true for any $N>0$ at any point x of a set E. Integrating these inequalities over the set E** we establish Property 3° for arbitrary nonnegative summable functions f_1 and f_2.

Proof of the remaining Properties 2° to 5° for such functions will be left to the reader.

We now proceed to discuss two other fundamental properties of arbitrary nonnegative summable functions.

Theorem 8.17 **(*complete additivity*).** *Let a set E be a sum of a countable number of mutually disjoint measurable sets E_k, i.e. let $E = \bigcup\limits_{k=1}^{\infty} E_k$. Then the following two statements are true.*

I. *If a nonnegative function $f(x)$ is summable on a set E, then $f(x)$ is summable on every set E_k, with*

$$\int_E f(x)\,dx = \sum_{k=1}^{\infty} \int_{E_k} f(x)\,dx. \tag{8.41}$$

II. *If a function $f(x)$ nonnegative on a set E is summable on every set E_k and if the series on the right of (8.41) converges, then $f(x)$ is summable on E too and equation (8.41) is true for it.*

Proof. (1) We first prove Theorems I and II for the *bounded* nonnegative integrable function $f(x)$. Let there be a constant M such that $f(x) \leqslant M$ everywhere on E. Set $R_n = \bigcup\limits_{k=n+1}^{\infty} E_k$ and notice that by Theorem 8.8 $|R_n| = \sum\limits_{k=n+1}^{\infty} |E_k| \to 0$ (as $n\to\infty$). But then by Properties 4°, 5°, and 1°

$$\int_E f(x)\,dx - \sum_{k=1}^{n} \int_{E_k} f(x)\,dx = \int_{R_n} f(x)\,dx \leqslant M \int_{R_n} dx = M|R_n| \to 0$$

* The constant α in Property 2° must be nonnegative here.

** We use Properties 5° and 3° for bounded integrable functions.

(as $n \to \infty$). The last relation proves Theorems I and II for the case of the bounded integrable function.

(2) We now prove Theorem I for an arbitrary nonnegative summable function. The summability of $f(x)$ on every E_k follows at once from the inequality $\int\limits_{E_k} (f)_N(x)\,dx \leqslant \int\limits_{E} (f)_N(x)\,dx$ and from the fact that the integral on the left of the inequality is nondecreasing in N. It remains to prove equation (8.41). Using what was proved in (1) above and the inequality $(f)_N(x) \leqslant f(x)$ we get

$$\int\limits_{E} (f)_N(x)\,dx = \sum_{k=1}^{\infty} \int\limits_{E_k} (f)_N(x)\,dx \leqslant \sum_{k=1}^{\infty} \int\limits_{E_k} f(x)\,dx. \tag{8.42}$$

Proceeding in the last inequality to the limit as $N \to \infty$ we have

$$\int\limits_{E} f(x)\,dx \leqslant \sum_{k=1}^{\infty} \int\limits_{E_k} f(x)\,dx. \tag{8.43}$$

On the other hand, by the properties proved in Section 8.4.3, for any m

$$\int\limits_{E} (f)_N(x)\,dx = \sum_{k=1}^{\infty} \int\limits_{E_k} (f)_N(x)\,dx \geqslant \sum_{k=1}^{m} \int\limits_{E_k} (f)_N(x)\,dx$$

and making in the last inequality first N approach ∞ and only then m approach ∞ we obtain the inequality

$$\int\limits_{E} f(x)\,dx \geqslant \sum_{k=1}^{\infty} \int\limits_{E_k} f(x)\,dx$$

which, in conjunction with (8.43), proves (8.41).

(3) We finally prove Theorem II for an arbitrary nonnegative summable function. Notice that it suffices to establish only the summability of $f(x)$ on a set E (for equation (8.41) will then follow from Theorem I we have already proved).

But summability of $f(x)$ on E follows at once from inequality (8.42) and from the convergence of the series on the right of the inequality. The proof of the theorem is complete.

Theorem 8.18 (**absolute continuity of the integral**). *If $f(x)$ is nonnegative and summable on a set E, then, for any positive ε there is a positive number δ such that whatever a measurable subset e of E with measure $|e|$ less than δ may be,*

$$\int\limits_{e} f(x)\,dx < \varepsilon.$$

Proof. (1) First let the nonnegative function $f(x)$ be *bounded*, i.e. let there be M such that $f(x) \leqslant M$. Then (by the properties established in Section 8.4.3)

$$\int_e f(x)\,dx \leqslant M \int_e dx = M|e| < M\delta < \varepsilon \text{ when } \delta < \frac{\varepsilon}{M}.$$

(2) We now prove the theorem for an arbitrary nonnegative summable function $f(x)$. On fixing an arbitrary $\varepsilon > 0$ we may choose (on the basis of the definition of summability) $N = N(\varepsilon)$ such that

$$\int_E [f(x) - (f)_N(x)]\,dx < \frac{\varepsilon}{2}. \tag{8.44}$$

Using (8.44) and the inequality $(f)_N(x) \leqslant N$ we get

$$\int_e f(x)\,dx = \int_e [f(x) - (f)_N(x)]\,dx + \int_e (f)_N(x) < \frac{\varepsilon}{2} + N \int_e dx =$$

$$= \frac{\varepsilon}{2} + N \cdot |e| < \frac{\varepsilon}{2} + N\delta < \varepsilon,$$

provided $\delta < \varepsilon/2N(\varepsilon)$. Thus the theorem is proved.

In conclusion we show two more properties true only for nonnegative summable functions.

Condition for a nonnegative summable function to be equivalent to zero: *If $f(x)$ is nonnegative, measurable and summable on a set E and if $\int_E f(x)\,dx$ is equal to zero, then $f(x)$ is equivalent to an identical zero on E.*

Proof. It suffices to prove that the measure of the set $E[f > 0]$ is equal to zero. We first show that for *any* $a > 0$ the measure of a set $E_a = E[f > a]$ is equal to zero. Indeed, if the measure $|E_a|$ were positive, we should obtain the inequality $\int_E f(x)\,dx \geqslant \int_{E_a} f(x)\,dx \geqslant$ $\geqslant aE_a > 0$ contradicting the condition $\int_E f(x)\,dx = 0$. Now it remains to notice that $E[f > 0] = \bigcup_{k=1}^{\infty} E[f > 1/k]$ from which it follows that $|E[f > 0]| \leqslant \sum_{k=1}^{\infty} |E[f > 1/k]| = 0$.

Majorizing criterion for the summability of a nonnegative measurable function: *If $f_1(x)$ is nonnegative and measurable on a set E and $f_2(x)$ is summable on E and if $f_1(x) \leqslant f_2(x)$ is true everywhere on E, then $f_1(x)$ is also summable on E.*

of either of the functions $f^+(x)$ and $f^-(x)$ and from the inequalities $f^+(x) \leqslant |f(x)|$, $f^-(x) \leqslant |f(x)|$ it follows, by virtue of the majorizing criterion for the summability of a nonnegative measurable function (see this at the end of Section 8.4.4), that either of the functions $f^+(x)$ and $f^-(x)$ is summable on E, and this exactly means that $f(x)$ is summable on E.

Thus, for the Lebesgue integral (in contrast to the Riemann integral) the integrability of $f(x)$ is equivalent to that of $|f(x)|$.

We now consider the *properties* of arbitrary summable functions.

Note immediately *the validity for arbitrary summable functions of Properties 2° to 5°* established for bounded integrable functions in Section 8.4.3 and for nonnegative summable functions in Section 8.4.4. That these properties are valid for arbitrary summable functions follows immediately from equation (8.45) and from the validity of the properties for nonnegative summable functions.

Finally, for arbitrary summable functions the properties of complete additivity and absolute continuity of the Lebesgue integral still hold (proof of these properties for nonnegative summable functions made the contents of Theorems 8.17 and 8.18 of Section 8.4.4). We give a formulation and some brief hints concerning the proof of these properties.

*Theorem 8.17** (*complete additivity*). *Let a set E be a sum of a countable number of mutually disjoint measurable sets E_k, i.e. let $E = \bigcup_{k=1}^{\infty} E_k$. Then the following two statements are true.*

I. *If $f(x)$ is summable on E, then $f(x)$ is summable on every set E_k, equation* (8.41) *being true.*

II. *If $f(x)$ is measurable and summable on every set E_k and if the series*

$$\sum_{k=1}^{\infty} \int_{E_k} |f(x)|\,dx$$

converges, then $f(x)$ is summable on E and equation (8.41) *is true.*

To prove Theorem I it suffices to apply Theorem 8.17, I, to the nonnegative functions $f^+(x)$ and $f^-(x)$ and use equation (8.45).

To prove Theorem II it suffices to consider that by Theorem 8.17, II, the function $|f(x)|$ is summable on E. But then so is $f(x)$ and equation (8.41) holds by virtue of Theorem I already proved.

*Theorem 8.18** (*absolute continuity of the integral*). *If $f(x)$ is summable on a set E, then for any positive ε there is a positive number δ such that, whatever the measurable subset e of E with measure $|e|$ less than δ, the inequality $\left|\int_e f(x)\,dx\right| < \varepsilon$ holds.*

To prove this it suffices to apply Theorem 8.18 to the nonnegative function $|f(x)|$ and use $\left|\int_E f(x)\,dx\right| \leqslant \int_E |f(x)|\,dx$.

8.4.6. Passage to the limit under the Lebesgue integral.

Definition. It is said that a sequence $\{f_n(x)\}$ of functions summable on a set E converges to a function $f(x)$ summable on the same set in $L(E)$ if

$$\lim_{n\to\infty} \int_E |f_n(x)-f(x)|\,dx = 0. \tag{8.46}$$

The convergence of $\{f_n(x)\}$ in $L(E)$ ensures the possibility of a term-by-term integration of $\{f_n(x)\}$ on E, for it follows from (8.46) that $\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx$.

Notice that *if a sequence $\{f_n(x)\}$ of functions measurable and summable on a set E converges in $L(E)$ to a function $f(x)$ measurable and summable on E, then $\{f_n(x)\}$ converges to $f(x)$ also in measure on E.*

Indeed, on fixing an arbitrary $\varepsilon > 0$ and denoting by E_n a set $E[|f-f_n| > \varepsilon]$ we have

$$\int_E |f_n(x)-f(x)|\,dx \geqslant \int_{E_n} |f_n(x)-f(x)|\,dx \geqslant \varepsilon |E_n|,$$

so that it follows from (8.46) that $|E_n| \to 0$ as $n \to \infty$.

Thus convergence in measure on E is weaker than convergence in $L(E)$ (and, as has already been established, weaker than convergence almost everywhere on E).

We prove, however, that under additional assumptions convergence in measure on E will imply convergence in $L(E)$.

Theorem 8.19 (Lebesgue's theorem). If a sequence $\{f_n(x)\}$ of functions measurable on a set E converges in measure on E to a function $f(x)$ measurable on E and if there is a function $F(x)$ summable on E such that for all n and almost all the points of E we have $|f_n(x)| \leqslant F(x)$, then $\{f_n(x)\}$ converges to $f(x)$ in $L(E)$.

Proof. We first show that the limit function $f(x)$ itself satisfies almost everywhere on E the inequality $|f(x)| \leqslant F(x)$. It follows from Theorem 8.15 that we can choose a subsequence $\{f_{n_k}(x)\}$ $(k = 1, 2, \ldots)$ of $\{f_n(x)\}$ converging to $f(x)$ almost everywhere on E. Proceeding in the inequality $|f_{n_k}(x)| \leqslant F(x)$ to the limit as $k \to \infty$ we get $|f(x)| \leqslant F(x)$ for almost all the points of E. From the inequality we have proved and from the majorizing criterion for the summability of a nonnegative measurable function (see this at the end of Section 8.4.4) it follows that $f(x)$ is *summable* on E.

On fixing an arbitrary $\varepsilon > 0$ and denoting by E_n a set $E[|f - f_n| > \varepsilon]$ we have*

$$\int_E |f(x) - f_n(x)|\,dx = \int_{E_n} |f(x) - f_n(x)|\,dx + \int_{E \setminus E_n} |f(x) - f_n(x)|\,dx \leqslant 2\int_{E_n} F(x)\,dx + \varepsilon |E|.$$

From this inequality and from the arbitrariness of $\varepsilon > 0$ it follows that to establish the convergence of $\{f_n(x)\}$ to $f(x)$ in $L(E)$ it suffices to prove that $\lim\limits_{n\to\infty} \int_{E_n} F(x)\,dx = 0$; but this follows immediately from Theorem 8.18* on the absolute continuity of the integral and from the fact that under the hypothesis $|E_n| \to 0$ as $n \to \infty$. Thus the theorem is proved.

Corollary (*Lebesgue's theorem on passage to the limit under the integral*). *If a sequence $\{f_n(x)\}$ of functions measurable on a set E converges almost everywhere on E to a limit function $f(x)$ and if there is a function $F(x)$ summable on E such that for all n and almost all the points of E we have $|f_n(x)| \leqslant F(x)$, then $f(x)$ is summable on E and*

$$\lim_{n\to\infty} \int_E f_n(x)\,dx = \int_E f(x)\,dx. \tag{8.47}$$

Proof. It follows from Theorem 8.13 that $f(x)$ is measurable on E. Hence it suffices to notice that convergence almost everywhere on E implies (by virtue of Theorem 8.14) convergence in measure on E; and to use Theorem 8.19.

Theorem 8.20 (*B. Levi's theorem*). *Let every function $f_n(x)$ be measurable and summable on a set E and let $f_n(x) \leqslant f_{n+1}(x)$ for all n and for almost all the points of E. Suppose further that there is a constant M such that $\left|\int_E f_n(x)\,dx\right| \leqslant M$ for all n. Then for almost every point x of E there is a finite limit $\lim\limits_{n\to\infty} f_n(x) = f(x)$, the limit function $f(x)$ being summable on E and equation* (8.47) *holding.*

Proof. We may assume without loss of generality that all $f_n(x)$ are *nonnegative* almost everywhere on E (otherwise we should take the nonnegative functions $g_n(x) = f_n(x) - f_1(x)$ instead of $f_n(x)$). Since $\{f_n(x)\}$ is nondecreasing almost everywhere on E, a limit function $f(x)$ is defined at almost all the points of E taking at them either finite values or values equal to $+\infty$. If we prove that limit function is summable on E, then it will follow that $f(x)$ has *finite* values almost everywhere on E, i.e. it will follow that $\{f_n(x)\}$ converges to $f(x)$ *almost everywhere* on E, and this and the inequality

* We consider that $|f_n(x) - f(x)| \leqslant 2F(x)$ almost everywhere on E.

$f_n(x) \leqslant f(x)$ (almost everywhere on E) will lead, by virtue of the corollary of the previous theorem, to equation (8.47).

So, to prove the theorem it suffices to establish that the limit function $f(x)$ is summable on E.

Notice that for any $N > 0$ the sequence* $\{(f_n)_N(x)\}$ converges to $(f)_N(x)$ almost everywhere on E, the bounded function $(f)_N(x)$ being summable on E and $(f_n)_N(x) \leqslant (f)_N(x)$ for all n and almost all points of E.

This ensures the applicability to $\{(f_n)_N(x)\}$ of the corollary of the previous theorem, by which

$$\lim_{n\to\infty} \int_E (f_n)_N(x)\,dx = \int_E (f)_N(x)\,dx.$$

From this and from the inequality**

$$\int_E f_n(x)\,dx \geqslant \int_E (f_n)_N(x)\,dx$$

we conclude that

$$\lim_{n\to\infty} \int_E f_n(x)\,dx \geqslant \int_E (f)_N(x)\,dx,$$

and since $\int_E f_n(x)\,dx \leqslant M$ for all n, so is

$$\int_E (f)_N(x)\,dx \leqslant M. \tag{8.48}$$

From inequality (8.48) and from the fact that the integral on the left of (8.48) is nondecreasing in N it follows that there is a limit

$$\lim_{N\to\infty} \int_E (f)_N(x)\,dx,$$

which just means that $f(x)$ is summable on E. Thus the theorem is proved.

We now formulate Theorem 8.20 in terms of a functional series (the theorem is widely used in this form).

If every function $u_n(x)$ is nonnegative almost everywhere on a set E, is measurable and summable on it, and if the series

$$\sum_{n=1}^{\infty} \int_E u_n(x)\,dx$$

---

* Recall that for any $N > 0$ and for any function $F(x)$ we set $(F)_N(x) = \min\{N, F(x)\}$.

** This inequality follows from $(f_n)_N(x) = \min\{N, f_n(x)\}$.

converges, then so does almost everywhere on E the series

$$\sum_{n=1}^{\infty} u_n(x), \tag{8.49}$$

the sum $S(x)$ of the series (8.49) *being summable on E and satisfying the condition*

$$\int_E S(x)\,dx = \sum_{n=1}^{\infty} \int_E u_n(x)\,dx.$$

Theorem 8.21 (Fatou's theorem). If a sequence $\{f_n(x)\}$ of functions measurable and summable on a set E converges almost everywhere on E to a limit function $f(x)$ and if there is a constant A such that $\int_E |f_n(x)|\,dx \leqslant A$ for all n, then $f(x)$ is summable on E and $\int_E |f(x)|\,dx \leqslant A$ for it.

Proof. We introduce into consideration functions $g_n(x) = \inf_{k \geqslant n} |f_k(x)|$* and notice that every function $g_n(x)$ is nonnegative and measurable** on E and that the sequence $\{g_n(x)\}$ is nondecreasing on E and converges to $|f(x)|$ for almost all points of E. In addition, for any n everywhere on E

$$g_n(x) \leqslant |f_n(x)|, \tag{8.50}$$

from which (by the majorizing criterion for the summability of a nonnegative measurable function, see this at the end of Section 8.4.4) it follows that $g_n(x)$ is summable on E. Applying to $\{g_n(x)\}$ Theorem 8.20 we get

$$\lim_{n\to\infty} \int_E g_n(x)\,dx = \int_E |f(x)|\,dx. \tag{8.51}$$

Since by (8.50) $\int_E g_n(x)\,dx \leqslant \int_E |f_n(x)|\,dx \leqslant A$ for any n, from (8.51) we get $\int_E |f(x)|\,dx \leqslant A$. Thus the theorem is proved.

8.4.7. Lebesgue classes $L^p(E)$. Recall that a linear space R is said to be *normalized* if the following two requirements hold: (1) there is a rule by which each element f of R is assigned a real number called the norm of that element and designated $\|f\|_R$, (2) the rule satisfies the following three axioms:

---

* The notation means that for every x the value of $g_n(x)$ is the infimum of $|f_n(x)|, |f_{n+1}(x)|, \ldots$.

** That $g_n(x)$ is measurable on E follows from Theorem 8.12.

1°. $\| f \|_R > 0$, if $f \neq 0$*, $\| f \|_R = 0$, if $f = 0$.

2°. $\| \lambda f \|_R = | \lambda | \cdot \| f \|_R$ for any element f and any real number λ.

3°. For any two elements f and g the so-called *triangle inequality* $\| f + g \|_R \leqslant \| f \|_R + \| g \|_R$ is true.

We shall consider in a normalized linear space R an arbitrary sequence of elements $\{f_n\}$.

Definition 1. A sequence $\{f_n\}$ of elements of a normalized linear space R is said to be **fundamental** *if*

$$\lim_{\substack{m \geqslant n \\ n \to \infty}} \| f_m - f_n \|_R = 0.$$

Definition 2. A sequence $\{f_n\}$ of elements of a normalized linear space R is said to **converge** *in R to an element f of R if*

$$\lim_{n \to \infty} \| f_n - f \|_R = 0.$$

Such convergence is sometimes called *convergence in norm* or *strong convergence* in R.

It is easy to prove that *any sequence of elements $\{f_n\}$ converging in R is always fundamental.* Indeed, if there is some element f such that $\| f_n - f \|_R \to 0$ as $n \to \infty$, then from the triangle inequality

$$\| f_m - f_n \|_R \leqslant \| f_m - f \|_R + \| f - f_n \|_R$$

it follows immediately that

$$\lim_{\substack{m \geqslant n \\ n \to \infty}} \| f_m - f_n \|_R = 0.$$

The question naturally arises as to whether any fundamental sequence of elements $\{f_n\}$ is convergent in R to some element f of R.

Definition 3. A normalized linear space R is said to be **complete** *if any fundamental sequence of elements $\{f_n\}$ of R converges in R to some element f of R.*

We shall discuss here an important class of normalized linear spaces due to Lebesgue and prove the completeness of these spaces.

Let a real number p satisfy the condition $p \geqslant 1$.

Definition 4. We shall say that a function $f(x)$ **belongs to the class** *(or space) $L^p(E)$ if $f(x)$ is measurable on a set E and $| f(x) |^p$ is summable on that set***.

It is easy to see that for any $p \geqslant 1$ the class $L^p(E)$ is a normalized linear space if the norm is introduced in it using the relation

$$\| f \|_{L^p(E)} = \| f \|_p = \left(\int_E | f(x) |^p \, dx \right)^{1/p}.$$

* 0 stands for the zero element of a linear space R.

** We shall not distinguish between functions equivalent on E, regarding them as a single element $L^p(E)$.

The linearity of such a space is obvious. It is easy to verify Axioms 1° to 3° using the definition of a normalized space. Axiom 1° follows immediately from the condition for a nonnegative summable function to be equivalent to zero (see this at the end of Section 8.4.4). Axiom 2° is quite obvious. Axiom 3° is obvious for $p = 1$, and for $p > 1$ it follows from the *Minkowski* inequality*

$$\left(\int_E |f(x)+g(x)|^p \, dx \right)^{1/p} \leqslant \left(\int_E |f(x)|^p \, dx \right)^{1/p} + \int_E |g(x)|^p \, dx \Big)^{1/p}$$

established in Supplement 1 to Chapter 10 of [1].

We now prove the following theorem**.

Theorem 8.22. For any $p \geqslant 1$ the space $L^p(E)$ is complete.

Proof. Let $\{f_n(x)\}$ be an arbitrary fundamental sequence of elements of the space $L^p(E)$. Set

$$\varepsilon_n = \sup_{m \geqslant n} \| f_m - f_n \|_p$$

(the supremum of $\| f_m - f_n \|_p$ is taken over the set of all m satisfying the inequality $m \geqslant n$). It follows from the condition for a sequence $\{f_n\}$ to be fundamental that $\varepsilon_n \to 0$ as $n \to \infty$. It follows that we can choose a subsequence of n_k ($k = 1, 2, \ldots$) such that the series***

$$\sum_{k=1}^{\infty} \varepsilon_{n_k} \tag{8.52}$$

converges.

From the Hölder**** inequality established in Supplement 1 to Chapter 10 of [1]

$$\left(\int_E |f(x) \cdot g(x)| \, dx \leqslant \left(\int_E |f(x)|^p \, dx \right)^{1/p} \cdot \int_E |g(x)^q \, dx \right)^{1/q}$$

$\left(p > 1, \ q = \frac{p}{p-1} \right)$ it follows when $p > 1$

---

* In that supplement the Minkowski inequality has been established for the case of the Riemann integral. In the case of the Lebesgue integral it suffices to establish this inequality only for bounded functions $f(x)$ and $g(x)$, and for such functions the proof is similar to that for the Riemann integral (it is sufficient to consider a *Lebesgue subdivision* of the set E).

** In a special form (relating to the so-called trigonometric system) this theorem was proved in 1907 by F. Riesz and independently by Fisher. In 1909 Hermann Weyl noticed that the relation to the trigonometric system is unessential and gave the more general formulation (for $p = 2$) presented here.

*** It suffices to take n_k such that $\varepsilon_{n_k} \leqslant 2^{-k}$.

**** In that supplement the Hölder inequality has been established for the Riemann integral. In the case of the Lebesgue integral it suffices to establish this inequality only for bounded functions $f(x)$ and $g(x)$, but for such functions the proof is similar to that for the Riemann integral (it is sufficient to consider a *Lebesgue subdivision* of a set E).

$$\int\limits_E |f_{n_{k+1}}(x)-f_{n_k}(x)|\,dx \leqslant \|f_{n_{k+1}}(x)-f_{n_k}(x)\|_p\cdot\Bigl(\int\limits_E 1^q\,dx\Bigr)^{1/q}\leqslant$$

$$\leqslant \varepsilon_{n_k}\cdot|E|^{\frac{p-1}{p}},$$

and from the last inequality and the convergence of the series (8.52) it follows that the series*

$$\sum_{k=1}^{\infty}\int\limits_E |f_{n_{k+1}}(x)-f_{n_k}(x)|\,dx \tag{8.53}$$

converges. From the convergence of (8.53) and from Theorem 8.20 (see the formulation of the theorem in terms of a series) we conclude that the series

$$\sum_{k=1}^{\infty}|f_{n_{k+1}}(x)-f_{n_{k+1}}(x)|$$

converges almost everywhere on E, and therefore so does the series

$$f_{n_1}(x)+\sum_{k=1}^{\infty}[f_{n_{k+1}}(x)-f_{n_k}(x)].$$

But this means that the kth subsum of that series equal to $f_{n_{k+1}}(x)$ converges almost everywhere on E to some function $f(x)$. Further, since $\|f_m(x)-f_{n_k}(x)\|_p\leqslant\varepsilon_m$ for any m and any $n_k\geqslant m$ and since $[f_m(x)-f_{n_k}(x)]\to[f_m(x)-f(x)]$ as $k\to\infty$ almost everywhere on E, by Theorem 8.21 (Fatou) $\|f_m(x)-f(x)\|_p\leqslant\varepsilon_m$ (for any m) and this just means that the sequence $\{f_m(x)\}$ converges in $L^p(E)$ to $f(x)$. Thus the theorem is proved.

8.4.8. **Concluding remarks.** The central point of the Lebesgue theory is *closure under the operation of proceeding to the limit* in the theory of measurable sets (Theorems 8.3 and 8.8), in the theory of measurable functions (Theorem 8.13), and in the theory of the integral (Theorem 8.22).

We discussed everything for the case of a single variable. In the case of n variables the scheme of constructing the theory remains the same, but instead of the interval (a, b) we should take an open n-dimensional parallelepiped $\prod\limits_{k=1}^{n}(a_k<x_k<b_k)$ (allowing $-\infty$ values for a_k and $+\infty$ values for b_k) to be the original (primary) set. In n dimensions the only qualitatively new feature of the theory is the so-called Fubini theorem on reduction of the n-fold multiple Lebesgue integral to an iterated integral of lower multiplicity. We shall skip this theorem.

* We need not use the Hölder inequality when $p = 1$, for the series (8.53) coincides with (8.52).

SUPPLEMENT 1

THE NECESSARY AND SUFFICIENT CONDITION FOR RIEMANN INTEGRABILITY

Without loss of generality, consider functions defined on a closed interval $[0, 1]$. For every such function $f(x)$ we introduce the so-called *Baire functions* $m(x)$ and $M(x)$ corresponding to the upper and lower limits of $f(x)$ at every point x*. So by definition

$$m(x) = \varliminf_{y \to x} f(y), \qquad M(x) = \varlimsup_{y \to x} f(y).$$

Notice that Baire functions may be defined in another way:

$$m(x) = \lim_{\delta \to 0+0} [\inf_{v_\delta(x)} f(y)], \qquad M(x) = \lim_{\delta \to 0+0} [\sup_{v_\delta(x)} f(y)],$$

where $v_\delta(x)$ is a δ-neighbourhood of a point x (in case x is an end point of $[0, 1]$ the right- or left-hand δ-half-neighbourhood of x respectively should be taken instead of the δ-neighbourhood).

Obviously *the function $f(x)$ is continuous at a point x_0 if and only if $f(x_0) = m(x_0) = M(x_0)$.*

Theorem 8.23. For a function $f(x)$ bounded on $[0, 1]$ to be Riemann integrable on $[0, 1]$ it is necessary and sufficient that $f(x)$ should be continuous almost everywhere on $[0, 1]$.

Proof. For any n, divide $[0, 1]$ into 2^n intervals $\Delta_k^{(n)} = \left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)$ $(k = 1, 2, 3, \ldots, 2^n)$ and introduce into consideration two step functions $\varphi_n(x)$ and $\Phi_n(x)$, setting them equal to $\inf\limits_{\Delta_k^{(n)}} f(y)$ and $\sup\limits_{\Delta_k^{(n)}} f(y)$ at every interval $\Delta_k^{(n)}$ and to zero at points $k/2^n$ $(k = 0, 1, \ldots, 2^n)$. Then on taking for every point $x \neq k/2^n$ a sequence of intervals $\Delta_k^{(n)}$ contracting to x we get

$$\lim_{n \to \infty} \varphi_n(x) = m(x), \qquad \lim_{n \to \infty} \Phi_n(x) = M(x). \tag{8.54}$$

The convergence (8.54) thus occurs *almost everywhere* on $[0, 1]$. Since the step functions $\varphi_n(x)$ and $\Phi_n(x)$ are a fortiori measurable on $[0, 1]$, it follows from (8.54) and Theorem 8.13 that so are the Baire functions $m(x)$ and $M(x)$.

From (8.54) we see that almost everywhere on $[0, 1]$

$$\lim_{n \to \infty} [\Phi_n(x) - \varphi_n(x)] = M(x) - m(x).$$

From the last relation it follows by virtue of the corollary of Theorem 8.19 that**

$$\lim_{n \to \infty} \int_0^1 [\Phi_n(x) - \varphi_n(x)]\, dx = \int_0^1 [M(x) - m(x)]\, dx. \tag{8.55}$$

---

* In case the function $f(x)$ is not bounded below (above) in an arbitrarily small neighbourhood of x, we set the lower (upper) limit of $f(x)$ equal to $-\infty$ $(+\infty)$ at that point.

** Henceforth all integrals in Supplement 1 are understood in the sense of Lebesgue.

It remains to notice that

$$\int_0^1 \Phi_n(x)\,dx = S_n, \qquad \int_0^1 \varphi_n(x)\,dx = s_n, \tag{8.56}$$

where S_n and s_n are respectively the upper and lower Darboux sums corresponding to the subdivision $\{\Delta_k^{(n)}\}$ $(k = 1, 2, 3, \ldots, 2^n)$.

From (8.55) and (8.56) it follows that

$$\lim_{n\to\infty}(S_n - s_n) = \int_0^1 [M(x) - m(x)]\,dx,$$

so that (by virtue of Chapter 10 of [1]) a necessary and sufficient condition for Riemann integrability reduces to the equation $\int_0^1 [M(x) - m(x)]\,dx = 0$. But the last equation, by virtue of the condition for a nonnegative measurable and summable function to be equivalent to zero (see Section 8.4.4), means that $M(x) - m(x) = 0$ almost everywhere on $[0, 1]$. Thus the theorem is proved.

SUPPLEMENT 2

THE NECESSARY AND SUFFICIENT CONDITION FOR THE LEBESGUE INTEGRABILITY OF A BOUNDED FUNCTION

Theorem 8.24. For a function $f(x)$ bounded on a measurable set E to be Lebesgue integrable on that set it is necessary and sufficient that the function should be measurable on E.

Proof. Proof of the sufficiency makes the content of Theorem 8.16, it is therefore necessary to prove only the necessity.

Let a function $f(x)$ be bounded and Lebesgue integrable on a measurable set E. This means that the upper and lower Lebesgue integrals of that function are equal and there is therefore a sequence of subdivisions $T_n = \{E_k^{(n)}\}$ of E such that the corresponding sequences of the upper sums $\{S_n\}$ and the lower sums $\{s_n\}$ satisfy the condition $S_n - s_n < 1/n$, each subsequent subdivision $T_n = \{E_k^n\}$ being a refinement of the previous subdivision $T_{n-1} = \{E_k^{n-1}\}$. (To construct such a sequence of subdivisions it suffices wherever necessary to take a product of subdivisions to be introduced).

Recall that by definition

$$S_n = \sum_k M_k^{(n)} \cdot |E_k^{(n)}|, \quad s_n = \sum_k m_k^{(n)} \cdot |E_k^{(n)}|,$$

where M_k^n and $m_k^{(n)}$ are respectively the supremum and infimum of $f(x)$ on a se $E_k^{(n)}$.

We define two sequences of functions $\{\overline{f}_n(x)\}$ and $\{\underline{f}_n(x)\}$ by setting $\overline{f}_n(x)$ equal to $M_k^{(n)}$ on $E_k^{(n)}$, and $\underline{f}_n(x)$ equal to $m_k^{(n)}$ on $E_k^{(n)}$.

It is obvious that for every n the two functions $\overline{f}_n(x)$ and $\underline{f}_n(x)$ are measurable on E (for these functions are linear combinations of characteristic functions of measurable sets $E_k^{(n)}$).

It is also obvious that the sequence $\{\overline{f}_n(x)\}$ is nonincreasing and $\{\underline{f}_n(x)\}$ is nondecreasing on E, with

$$\underline{f}_n(x) \leqslant f(x) \leqslant \overline{f}_n(x) \tag{8.57}$$

for any n at each point of E. Set $\overline{f}(x) = \lim\limits_{n\to\infty} \overline{f}_n(x)$, $\underline{f}(x) = \lim\limits_{n\to\infty} \underline{f}_n(x)$. From (8.57) we conclude that

$$\underline{f}(x) \leqslant f(x) \leqslant \overline{f}(x), \tag{8.58}$$

$\overline{f}(x)$ and $\underline{f}(x)$ being measurable on E by virtue of Theorem 8.13.

From Theorem 8.20 (B. Levi) we get

$$\lim_{n\to\infty} \int\limits_E [\overline{f}_n(x) - \underline{f}_n(x)]\,dx = \int\limits_E [\overline{f}(x) - \underline{f}(x)]\,dx. \tag{8.59}$$

From the definition of the functions $\overline{f}_n(x)$ and $\underline{f}_n(x)$ it follows that $\int\limits_E [\overline{f}_n(x) - \underline{f}_n(x)]\,dx = S_n - s_n$, with $\lim\limits_{n\to\infty} (S_n - s_n) = 0$ by construction. By virtue of (8.59) this leads to the equation $\int\limits_E [\overline{f}(x) - \underline{f}(x)]\,dx = 0$.

From the last equation and from the nonnegativity and measurability of the function $[\overline{f}(x) - \underline{f}(x)]$ it follows by virtue of Section 8.4.4 that $\overline{f}(x) - \underline{f}(x) = 0$ almost everywhere on E. Consequently, by (8.58) $\underline{f}(x) = f(x) = \overline{f}(x)$ almost everywhere on E, and since $\underline{f}(x)$ and $\overline{f}(x)$ are measurable on E, by Property 4° of Section 8.3.2 so is $f(x)$. Thus the theorem is proved.

CHAPTER 9

INTEGRALS DEPENDENT ON PARAMETERS

In this chapter we study a special class of functions characterized by a common name of "parameter-dependent integrals". We may get an idea of these functions if we integrate a function of two variables x and y with respect to x with every y fixed. As a result we shall obviously obtain a function dependent on the parameter y.

The questions naturally arise as to whether such functions are continuous, integrable and differentiable. These questions are to be studied in the present chapter.

It is quite clear that integration with respect to the independent variable x must not necessarily be proper—if the domain of the function $f(x, y)$ is an infinite band $\Pi = \{a \leqslant x < \infty,\ c \leqslant y \leqslant d\}$, then integration with respect to x, with y fixed, is carried out over a half-line and therefore the corresponding integral over the variable x is improper. This leads to the concept of parameter-dependent improper integrals. In this chapter we shall study the properties of such integrals.

We stress that throughout the chapter we discuss Riemann, and not Lebesgue, integrable functions and that all integrals, whether proper or improper, are understood in the Riemann sense.

9.1. PROPER INTEGRALS DEPENDENT ON A PARAMETER

9.1.1. The parameter-dependent integral. Suppose in a rectangle $\Pi = \{a \leqslant x \leqslant b,\ c \leqslant y \leqslant d\}$ a function $f(x, y)$ integrable over x on the closed interval $a \leqslant x \leqslant b$ is defined for any fixed y in the closed interval $c \leqslant y \leqslant d$. In this case on $c \leqslant y \leqslant d$ the function

$$I(y) = \int_a^b f(x, y)\, dx \tag{9.1}$$

is defined called the *integral dependent on a parameter* y. The function $f(x, y)$ may be given on a set of a more general form as well. For example, the set $D = \{a(y) \leqslant x \leqslant b(y),\ c \leqslant y \leqslant d\}$ may serve as the domain of $f(x, y)$. In this case the function of y is defined on $[c, d]$ using relations (9.1) but the limits of integration, a and b,

will depend on y. We shall first study the case where the limits of integration are constant.

9.1.2. Continuity, integrability and differentiability of parameter-dependent integrals. The following theorems give the answer to the above questions. In these theorems Π will denote the rectangle $\{a \leqslant x \leqslant b,\ c \leqslant y \leqslant d\}$.

Theorem 9.1. If a function $f(x, y)$ is continuous in Π, then the function $I(y)$ defined by relation (9.1) is continuous on $[c, d]$.

Proof. It follows from formula (9.1) that the increment $\Delta I = I(y + \Delta y) - I(y)$ of $I(y)$ equals

$$\Delta I = \int_a^b [f(x, y+\Delta y) - f(x, y)]\, dx. \tag{9.2}$$

Since by the Cantor theorem $f(x, y)$ is uniformly continuous in Π, given $\varepsilon > 0$ we can find $\delta > 0$ such that for all x in $[a, b]$ and all y and $(y + \Delta y)$ in $[c, d]$ such that $|\Delta y| < \delta$ we have $|f(x, y+\Delta y) - f(x, y)| < \varepsilon/(b - a)$. But then it follows from relation (9.2) that when $|\Delta y| < \delta$ we have the inequality $|\Delta I| < \varepsilon$ which implies that $I(y)$ is continuous at each point y of $[c, d]$. Thus the theorem is proved.

Theorem 9.2. If $f(x, y)$ is continuous in Π, then $I(y)$ is integrable on $[c, d]$. Moreover,

$$\int_c^d I(y)\, dy = \int_c^d \left[\int_a^b f(x, y)\, dx\right] dy = \int_a^b dx \int_c^d f(x, y)\, dy. \tag{9.3}$$

In other words, under the hypotheses of the theorem *a parameter-dependent integral may be integrated with respect to the parameter under the integral.*

Proof. By Theorem 9.1 $I(y)$ is continuous on $[c, d]$ and is therefore integrable on that interval. The validity of formula (9.3) follows from the equality of the iterated integrals occurring in (9.3) (they are equal to the double integral $\iint\limits_{\Pi} f(x, y)\, dx\, dy$). Thus the theorem is proved

Remark. In (9.3) instead of the upper limit d of integration with respect to y we may put any number from the closed interval $[c, d]$.

Theorem 9.3. If $f(x, y)$ and its partial derivative $\frac{\partial f}{\partial y}$ are continuous in Π, then $I(y)$ is differentiable on $[c, d]$ and its derivative $\frac{dI}{dy}$ can be found from the formula

$$\frac{dI}{dy} = \int_a^b \frac{\partial f(x, y)}{\partial y}\, dx. \tag{9.4}$$

In other words, under the hypotheses of the theorem *a parameter-dependent integral may be differentiated with respect to the parameter under the integral.*

Proof. Consider the following auxiliary function:

$$g(y) = \int_a^b \frac{\partial f(x, y)}{\partial y}\, dx. \tag{9.5}$$

Since $\frac{\partial f}{\partial y}$ is continuous in Π, by Theorem 9.1 $g(y)$ is continuous on $[c, d]$ and the integral of the function over $[c, y]$ can be found from the formula for integration under the integral. According to the remark to Theorem 9.2 we get

$$\int_c^y g(t)\, dt = \int_a^b dx \int_c^y \frac{\partial f(x, t)}{\partial t}\, dt = \int_a^b f(x, y)\, dx - \int_a^b f(x, c)\, dx. \tag{9.6}$$

Since $\int_a^b f(x, y)\, dx = I(y)$ and $\int_a^b f(x, c)\, dx = I(c)$, from relation (9.6) we obtain the following representation for $I(y)$:

$$I(y) = \int_c^y g(t)\, dt + I(c). \tag{9.7}$$

As is known, the derivative of an integral with a variable upper limit of a continuous function $g(t)$ exists and is equal to the value of that function at a point y. Therefore the function $I(y)$ is differentiable and its derivative $\frac{dI}{dy}$ is equal to $g(y)$. Turning to formula (9.5) for $g(y)$ we see that relation (9.4) is true. Thus the theorem is proved.

9.1.3. The case where the limits of integration depend on a parameter. We have already said that a case is possible where the limits of integration depend on a parameter. We shall assume that the function $f(x, y)$ is given in a rectangle Π containing a domain D defined by the relations $\{a(y) \leqslant x \leqslant b(y),\ c \leqslant y \leqslant d\}$ (Fig. 9.1). If for any fixed y in $[c, d]$ the function $f(x, y)$ is integrable over x on $[a(y)], [b(y)]$, then obviously the following function is defined on $[c, d]$:

$$I(y) = \int_{a(y)}^{b(y)} f(x, y)\, dx \tag{9.8}$$

which is *an integral dependent on the parameter whose limits of integration are also parameter-dependent.*

We shall investigate such integrals for continuity and differentiability with respect to the parameter. The following theorems give the answer to the above questions.

Theorem 9.4. Let $f(x, y)$ be continuous on Π and let $a(y)$ and $b(y)$ be continuous on $[c, d]$. Then $I(y)$ defined by relation (9.8) is continuous on $[c, d]$.

Proof. Choose an arbitrary y_0 in $[c, d]$ and represent $I(y)$ as

$$I(y) = \int_{a(y_0)}^{b(y_0)} f(x, y)\,dx + \int_{b(y_0)}^{b(y)} f(x, y)\,dx - \int_{a(y_0)}^{a(y)} f(x, y)\,dx. \tag{9.9}$$

Since the first integral on the right of (9.9) is an integral dependent on the parameter y, with constant limits of integration and constant

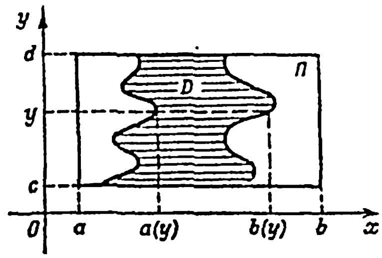


Fig. 9.1

integrand, it is a continuous function of y and tends therefore to $I(y_0)$ as $y \to y_0$. For the other two integrals we obtain the following estimates:

$$\left| \int_{b(y_0)}^{b(y)} f(x, y)\,dx \right| \leqslant M\,|b(y) - b(y_0)|,$$

$$\left| \int_{a(y_0)}^{a(y)} f(x, y)\,dx \right| \leqslant M\,|a(y) - a(y_0)|,$$

where $M = \sup\limits_{\Pi} |f(x, y)|$. From the last inequalities and from the continuity of the functions $a(y)$ and $b(y)$ it follows that as $y \to y_0$ the last two integrals on the right of (9.9) tend to zero. Thus the limit on the right of (9.9) as $y \to y_0$ exists and equals $I(y_0)$. So $I(y)$ is continuous at any point y_0 of $[c, d]$, i.e. is continuous on $[c, d]$. Thus the theorem is proved.

We now prove the theorem on the differentiability of the integral $I(y)$ with respect to the parameter.

Theorem 9.5. Let $f(x, y)$ and its derivative $\frac{\partial f}{\partial y}$ be continuous in Π. Also let $a(y)$ and $b(y)$ be differentiable on $[c, d]$. Then the function

$I(y)$ defined by relation (9.8) is differentiable on $[c, d]$ and its derivative $I'(y)$ is given by the formula

$$I'(y)=\int_{a(y)}^{b(y)} \frac{\partial f}{\partial y}\,dx+b'(y)\,f(b(y),\ y)-a'(y)\,f(a(y),\ y). \tag{9.10}$$

Proof. Choose an arbitrary y_0 in $[c, d]$ and represent $I(y)$ in the form (9.9). The first integral on the right of (9.9) is an integral dependent on the parameter y, with constant limits of integration. Since under the hypothesis $f(x, y)$ and $\frac{\partial f}{\partial y}$ are continuous in Π, by Theorem 9.3 the first term is a differentiable function at a point y_0 and the derivative of that function at that point is equal to $\int_{a(y_0)}^{b(y_0)} \frac{\partial f(x, y_0)}{\partial y}\,dx$. We prove that the second term on the right of (9.9) has a derivative at the point y_0. Since that second term vanishes at $y=y_0$, is suffices to show the existence of the following limit:

$$\lim_{y\to y_0} \frac{\int_{b(y_0)}^{b(y)} f(x, y)\,dx}{y-y_0} \tag{9.11}$$

which by definition is just equal to the desired derivative.

We transform the integral in the numerator of formula (9.11). From the mean value formula we have

$$\int_{b(y_0)}^{b(y)} f(x, y)\,dx=f(\bar{x}, y)\,(b(y)-b(y_0)), \tag{9.12}$$

with $\bar{x}$ contained between $b(y_0)$ and $b(y)$. Substituting the expression for the integral from formula (9.12) into the numerator of the expression (9.11) and considering that by continuity $f(\bar{x}, y)\to f(b(y_0), y_0)$ as $y\to y_0$ and $\frac{b(y)-b(y_0)}{y-y_0}\to b'(y_0)$ as $y\to y_0$ we see that the limit (9.11) of interest to us exists and equals $b'(y_0)\,f(b(y_0), y_0)$. Reasoning quite similarly we see that the third term on the right of (9.9) has also a derivative at the point y_0 equal to $a'(y_0)\,f(a(y_0), y_0)$.

So we have proved that $I(y)$ is differentiable at an arbitrary point y_0 of $[c, d]$ and that its derivative $I'(y_0)$ is given by formula (9.10). Thus the theorem is proved.

Remark. Theorems 9.4 and 9.5 are also true in the case where $f(x, y)$ is given in a domain D alone and satisfies in that domain the same requirements as in Π.

9.2. IMPROPER INTEGRAL DEPENDENT ON A PARAMETER

9.2.1. The parameter-dependent improper integral of the first kind. The uniform convergence of a parameter-dependent improper integral. The symbol Π_∞ will denote a half-band $\{a \leqslant x < \infty,\ c \leqslant y \leqslant d\}$.

Suppose in Π_∞ a function integrable over x in the improper sense on the half-line $a \leqslant x < \infty$ is given, with any y of $[c, d]$ fixed. Under these conditions the following function is defined on $[c, d]$:

$$I(y) = \int_a^\infty f(x, y)\, dx \tag{9.13}$$

which is called *an improper integral of the first kind dependent on the parameter* y. The integral (9.13) is said *to converge* on the closed interval $[c, d]$.

In the theory of parameter-dependent improper integrals an important role is played by the concept of *uniform convergence*. We shall formulate this notion.

Definition. An improper integral (9.13) *is said to be* **uniformly convergent** *with respect to a parameter* y *on a closed interval* $[c, d]$ *if it converges on* $[c, d]$ *and if given any* $\varepsilon > 0$ *we can find* $A \geqslant a$ *dependent only on* ε *such that for any* $R > A$ *and for every* y *in* $[c, d]$

$$\left| \int_R^\infty f(x, y)\, dx \right| < \varepsilon. \tag{9.14}$$

We formulate the Cauchy criterion for the uniform convergence of parameter-dependent improper integrals.

Theorem 9.6. For an improper integral (9.13) *to converge uniformly with respect to a parameter* y *on a closed interval* $[c, d]$ *it is necessary and sufficient that given any* $\varepsilon > 0$ *we should be able to find a number* $A \geqslant a$ *dependent only on* ε *such that for any* R' *and* R'' *greater than* A *and for every* y *in* $[c, d]$

$$\left| \int_{R'}^{R''} f(x, y)\, dx \right| < \varepsilon.$$

The validity of this criterion follows immediately from the definition of uniform convergence.

For applications it is appropriate to point out a number of sufficient tests for the uniform convergence of parameter-dependent improper integrals.

Theorem 9.7 (Weierstrass test). Let a function $f(x, y)$ *be defined in* Π_∞ *and integrable over* x *on any closed interval* $[a, R]$ *for every* y *in* $[c, d]$. *Suppose further that for all the points of* Π_∞

Then the convergence of $\int_a^{\infty} g(x)\,dx$ *implies the uniform convergence of the integral* (9.13) *with respect to* y *on* $[c, d]$.

Proof. By the Cauchy criterion for the convergence of the integral of $g(x)$ (see Theorem 3.1) given any $\varepsilon > 0$ we can find $A \geqslant a$ such that for all $R'' > R' \geqslant A$

$$\int_{R'}^{R''} g(x)\,dx < \varepsilon.$$

Applying inequality (9.15) we get

$$\left|\int_{R'}^{R''} f(x, y)\,dx\right| \leqslant \int_{R'}^{R''} g(x)\,dx < \varepsilon$$

for every y in $[c, d]$.

This just means that the Cauchy criterion for the uniform convergence of the integral (9.13) holds.

Corollary. Let a function $\varphi(x, y)$ *defined in a half-band* Π_∞ *be bounded in that half-band and integrable over* x *on any interval* $[a, R]$ *for each* $y \in [c, d]$. *Then if the integral*

$$\int_a^{\infty} |h(x)|\,dx$$

converges, then so does uniformly with respect to y *on* $[c, d]$ *the integral*

$$\int_a^{\infty} \varphi(x, y)\,h(x)\,dx.$$

To prove this it suffices to put in Theorem 9.7

$$f(x, y) = \varphi(x, y)\,h(x), \quad g(x) = M\,|h(x)|, \text{ where } M = \sup_{\Pi_\infty} |\varphi(x, y)|.$$

Note that the Weierstrass test is a sufficient test for the uniform convergence of parameter-dependent improper integrals which guarantees their absolute convergence. As in the proof of Theorem 3.4 we can establish the following sufficient test for uniform convergence that is applicable to conditionally convergent integrals as well. The following *statement* (*Abel-Dirichlet test*) is true.

Let a function $f(x, y)$ *be defined in a half-band* Π_∞, *be integrable over* x *on any closed interval* $[a, R]$ *for each* $y \in [c, d]$, *and satisfy with some constant* $M > 0$ *the condition*

$$\left|\int_a^{x} f(t, y)\,dt\right| \leqslant M.$$

Also suppose that the function $g(x)$ defined for $x \geqslant a$ and monotone nonincreasing tends to zero as $x \to +\infty$. Then the improper integral

$$\int_a^\infty f(x, y)\, g(x)\, dx$$

converges uniformly with respect to y on $[c, d]$.

The next test for uniform convergence relates to integrals of nonnegative functions.

Theorem 9.8 (Dini test). Let $f(x, y)$ be a function continuous and nonnegative in a half-band Π_∞ and suppose for each $y \in [c, d]$ the improper integral

$$I(y) = \int_a^\infty f(x, y)\, dx \tag{9.13}$$

converges.

Suppose further that $I(y)$ is continuous on $[c, d]$. Then the integral (9.13) converges uniformly with respect to y on $[c, d]$.

Proof. Consider the sequence of functions

$$I_n(y) = \int_a^{a+n} f(x, y)\, dx,$$

each by virtue of Theorem 9.1 continuous on $[c, d]$. Since the integrand $f(x, y)$ is nonnegative, $I_n(y)$, being monotone nondecreasing, converges on $[c, d]$ to the continuous function $I(y)$. Consequently, by Theorem 1.5 (Dini test for functional sequences) the sequence $I_n(y)$ converges to $I(y)$ uniformly on $[c, d]$. This means that given any $\varepsilon > 0$ we can find N such that

$$I(y) - I_N(y) = \int_{a+N}^\infty f(x, y)\, dx < \varepsilon$$

for all y of $[c, d]$ at once. From the nonnegativity of $f(x, y)$ it follows that for any $R \geqslant N + a$ and any $y \in [c, d]$

$$0 \leqslant \int_R^\infty f(x, y)\, dx < \varepsilon$$

This just means that the integral (9.13) is uniformly convergent. Thus the theorem is proved.

9.2.2. Continuity, integrability and differentiability of parameter-dependent improper integrals. The following two theorems are true.

Theorem 9.9. Let $f(x, y)$ be continuous in Π_∞ and let the integral (9.13) converge uniformly on a closed interval $[c, d]$. Then that integral is a continuous function of y on $[c, d]$.

Proof. Consider the sequence of functions

$$I_n(y) = \int_a^{a+n} f(x, y)\,dx,$$

each by virtue of Theorem 9.1 continuous on $[c, d]$. Obviously the uniform convergence of the integral (9.13) implies the uniform convergence to $I(y)$ of the functional sequence $I_n(y)$. In such a case the continuity of $I(y)$ follows from Theorem 1.7.

Theorem 9.10. Let $f(x, y)$ *and its partial derivative* $\frac{\partial f}{\partial y}$ *be continuous in* Π_∞. *Suppose further that for some* y *in* $[c, d]$ *the integral* $I(y) = \int_a^\infty f(x, y)\,dx$ *converges and* $\int_a^\infty \frac{\partial f}{\partial y}\,dx$ *converges uniformly with respect to* y *on* $[c, d]$. *Under these conditions* $I(y)$ *is differentiable on* $[c, d]$ *and its derivative* $I'(y)$ *can be found from the formula*

$$I'(y) = \int_a^\infty \frac{\partial f}{\partial y}\,dx. \tag{9.16}$$

In other words, under the hypotheses of the theorem *differentiation with respect to a parameter can be carried out under the parameter-dependent improper integral.*

Proof. Consider the sequence of functions

$$I_n(y) = \int_a^{a+n} f(x, y)\,dx.$$

By Theorem 9.3 each of the functions $I_n(y)$ is differentiable on $[c, d]$ and

$$I'_n(y) = \int_a^{a+n} \frac{\partial f(x, y)}{\partial y}\,dx. \tag{9.17}$$

From the hypotheses of the theorem it follows that the sequence of the integrals on the right of (9.17) converges uniformly on $[c, d]$. Consequently, uniformly converging to the same limit function is the sequence of the derivatives $I'_n(y)$. Applying Theorem 1.9 we obtain equation (9.16).

We prove a theorem on *proper integration of the parameter-dependent improper integral.*

Theorem 9.11. If the hypotheses of Theorem 9.9 *hold, then the integral* (9.13) *can be integrated with respect to the parameter* y *on*

a closed interval $[c, d]$, *with*

$$\int_c^d I(y)\,dy = \int_c^d dy \int_a^\infty f(x, y)\,dx = \int_a^\infty dx \int_c^d f(x, y)\,dy. \tag{9.18}$$

In other words, under the hypotheses of the theorem *a parameter-dependent improper integral can be integrated with respect to the parameter under the improper integral.*

Proof. Since the hypotheses of Theorem 9.9 hold, the function $I(y)$ is continuous on $[c, d]$ and is therefore integrable on $[c, d]$. We proceed to prove relations (9.18).

Using the property of uniform convergence of the integral (9.13) we can, for a given $\varepsilon > 0$, find $A \geqslant a$ such that when $R \geqslant A$ for every y in $[c, d]$

$$\left|\int_R^\infty f(x, y)\,dx\right| < \left|\frac{\varepsilon}{d-c}\right|. \tag{9.19}$$

Assuming further $R \geqslant A$ and using the possibility of inverting the order of integration for parameter-dependent improper integrals we turn to the following obvious equations:

$$\int_c^d I(y)\,dy = \int_c^d \left[\int_a^R f(x, y)\,dx + \int_R^\infty f(x, y)\,dx\right] dy =$$
$$= \int_a^R dx \left[\int_c^d f(x, y)\,dy\right] + \int_c^d dy \left[\int_R^\infty f(x, y)\,dx\right].$$

From these and inequality (9.19) it follows that for all $R \geqslant A$

$$\left|\int_c^d I(y)\,dy - \int_a^R dx\right| \left[\int_c^d f(x, y)\,dy\right] < \varepsilon$$

which means that the improper integral $\int_a^\infty dx \int_c^d f(x, y)\,dy$ over the variable x converges and equals the number $\int_c^d I(y)\,dy$. Thus the theorem is proved.

Remark. Obviously, in relation (9.18) we may substitute for the upper limit d of integration any number of the closed interval $[c, d]$.

Corollary. If $f(x, y)$ *is continuous and nonnegative in* Π_∞ *and the integral* (9.13) *is a continuous function on* $[c, d]$, *then formula* (9.18) *is true.*

Indeed, under the requirements formulated all the conditions of the Dini test for the uniform convergence of the integral (9.13) hold (see Theorem 9.8). The statement of the corollary is thus true.

We now prove a theorem on *improper integration of the parameter-dependent improper integral.*

Theorem 9.12. Let $f(x, y)$ be continuous and nonnegative when $x \geqslant a$ and $y \geqslant c$. Also let the integrals

$$I(y) = \int_a^\infty f(x, y)\,dx \text{ and } K(x) = \int_c^\infty f(x, y)\,dy$$

be continuous for $y \geqslant c$ and $x \geqslant a$ respectively. Then the convergence of one of the following two improper integrals

$$\int_c^\infty I(y)\,dy = \int_c^\infty dy \int_a^\infty f(x, y)\,dx \text{ and } \int_a^\infty K(x)\,dx = \int_a^\infty dx \int_c^\infty f(x, y)\,dy$$

implies the convergence of the other and equality of the two integrals.

Proof. Assume that the integral $\int_c^\infty I(y)\,dy$ converges. We must prove that the integral $\int_a^\infty K(x)\,dx$ converges and equals $\int_c^\infty I(y)\,dy$. In other words, it is necessary to prove that given any $\varepsilon > 0$ we can find $A \geqslant a$ such that for $\overline{R} \geqslant A$

$$\left| \int_0^\infty I(y)\,dy - \int_a^R K(x)\,dx \right| < \varepsilon. \tag{9.20}$$

It follows from the hypotheses of the theorem that given any fixed $\overline{R} \geqslant a$ the hypotheses of the corollary of Theorem 9.11 hold for $f(x, y)$ in the half-band $\{a \leqslant x \leqslant \overline{R},\ c \leqslant y < \infty\}$. For any $\overline{R} \geqslant a$ therefore

$$\int_a^{\overline{R}} K(x)\,dx = \int_a^{\overline{R}} dx \int_c^\infty f(x, y)\,dy = \int_c^\infty dy \int_a^{\overline{R}} f(x, y)\,dx.$$

Using these equations and the convergence of the integral $\int_c^\infty I(y)\,dy$ we transform the difference under the modulus sign in inequality

(9.20). For any $\bar{\bar{R}}$ greater than c we write the equation

$$\int_c^\infty I(y)\,dy - \int_a^{\bar{R}} K(x)\,dx = \int_c^\infty dy \int_a^\infty f(x, y)\,dx - \int_c^\infty dy \int_a^{\bar{R}} f(x, y)\,dx =$$

$$= \int_c^\infty dy \int_{\bar{R}}^\infty f(x, y)\,dx = \int_{\bar{\bar{R}}}^\infty dy \int_{\bar{R}}^\infty f(x, y)\,dx + \int_c^{\bar{\bar{R}}} dy \int_{\bar{R}}^\infty f(x, y)\,dx. \tag{9.21}$$

We proceed to evaluate the last integrals in relation (9.21). Since under the hypothesis $\int_c^\infty I(y)\,dy$ converges, given $\varepsilon > 0$, we can find $\bar{\bar{R}} > c$ such that $0 \leqslant \int_{\bar{\bar{R}}}^\infty I(y)\,dy < \varepsilon/2$*. We replace $I(y)$ in these inequalities by its expression as an integral to obtain the following inequalities: $0 \leqslant \int_{\bar{\bar{R}}}^\infty dy \int_a^\infty f(x, y)\,dx < \varepsilon/2$. From these and from the nonnegativity of $f(x, y)$ we conclude that for a chosen $\bar{\bar{R}} > c$ and any $\bar{R} \geqslant a$

$$0 \leqslant \int_{\bar{\bar{R}}}^\infty dy \int_{\bar{R}}^\infty f(x, y)\,dx < \varepsilon/2. \tag{9.22}$$

We now fix $\bar{\bar{R}}$ as shown above and use the arbitrariness of the choice of $\bar{R}$. In the half-band $\{a \leqslant x < \infty,\ c \leqslant y \leqslant \bar{\bar{R}}\}$ the function $f(x, y)$ satisfies all the conditions of the Dini test for the uniform convergence of improper integrals (see Theorem 9.8). Given $\varepsilon > 0$ therefore we can choose $A \geqslant a$ in such a way that for any $\bar{R} \geqslant A$ and for every y in a closed interval $[c, \bar{\bar{R}}]$ we have $0 \leqslant \int_{\bar{R}}^\infty f(x, y)\,dx < \varepsilon/2\,(r - c)$** from which we obtain the following estimate:

$$0 \leqslant \int_c^{\bar{\bar{R}}} dy \int_{\bar{R}}^\infty f(x, y)\,dx < \varepsilon/2. \tag{9.23}$$

---

* The left-hand one of these inequalities follows from the nonnegativity of $f(x, y)$ when $x \geqslant a$ and $y \geqslant c$.

** The left-hand one of these inequalities follows from the nonnegativity of $f(x, y)$ when $x \geqslant a$ and $y \geqslant c$.

Turning to the expression (9.21) and estimates (9.22) and (9.23) of the last integrals in (9.21) we see that given an arbitrary $\varepsilon > 0$ we can choose $A \geqslant a$ in such a way that for any $\bar{R} \geqslant A$ we have inequality (9.20). This completes the proof of the theorem.

9.2.3. Parameter-dependent improper integrals of the second kind. We introduce the concept of parameter-dependent improper integrals of the second kind. Let $f(x, y)$ be a function given in a half-open rectangle $\Pi = \{a \leqslant x < b,\ c \leqslant y \leqslant d\}$. Assum that for any fixed y in $[c, d]$ the improper integral of the second kind $\int_a^b f(x, y)\,dx$ converges. Under these conditions the function

$$I(y) = \int_a^b f(x, y)\,dx \tag{9.24}$$

is defined on $[c, d]$ called *an improper integral of the second kind dependent on the parameter y.*

In the theory of such integrals an important role is played by the concept of *uniform convergence*. We shall formulate this notion.

Definition. An improper integral (9.24) *is said to be* **uniformly convergent** *with respect to parameter y on a closed interval $[c, d]$ if it converges for every y in $[c, d]$ and given any $\varepsilon > 0$ we can find $\delta > 0$ dependent only on ε such that for any α in the interval $0 < \alpha < \delta$ and for every y in $[c, d]$*

$$\left| \int_{b-\alpha}^{b} f(x, y)\,dx \right| < \varepsilon.$$

For improper integrals of the second kind it is easy to formulate and prove theorems on continuity, integrability and differentiability with respect to a parameter.

Note that using the transformations of the variable x in Section 3.2.2 improper integrals of the second kind dependent of the parameter y can be reduced to parameter-dependent improper integrals of the first kind.

9.3. APPLICATION OF THE THEORY OF PARAMETER-DEPENDENT INTEGRALS TO THE EVALUATION OF IMPROPER INTEGRALS

The operations on parameter-dependent improper integrals substantiated in the preceding section allow computation of various improper integrals.

Consider examples of computing and investigating the properties of such integrals.

1°. Prove that the integral

$$I(\alpha)=\int_0^{\infty} e^{-\alpha x}\frac{\sin x}{x}\,dx \tag{9.25}$$

whose integrand at a point $x=0$ is by definition equal to unity converges uniformly with respect to α on a half-line $0\leqslant\alpha<\infty$. We first obtain some estimates. First note that

$$\int e^{-\alpha x}\sin x\,dx=-\frac{e^{-\alpha x}(\alpha\sin x+\cos x)}{1+\alpha^2}+C=\Phi(\alpha,\ x)+C.$$

Clearly, for $\alpha\geqslant 0$ and $x\geqslant 0$ the function $\Phi(\alpha,\ x)$ (antiderivative for the function $e^{-\alpha x}\sin x$) is bounded:

$$|\Phi(\alpha,\ x)|\leqslant\frac{1+\alpha}{1+\alpha^2}\leqslant 2. \tag{9.26}$$

Evaluate the following integral:

$$\int_R^{\infty} e^{-\alpha x}\frac{\sin x}{x}\,dx \qquad (R>0).$$

Integrating by parts for any fixed $\alpha\geqslant 0$ we find

$$\left|\int_R^{\infty} e^{-\alpha x}\frac{\sin x}{x}\,dx\right|=\left|\left[\frac{\Phi(\alpha,\ x)}{x}\right]_R^{\infty}+\int_R^{\infty}\frac{\Phi(\alpha,\ x)}{x^2}\,dx\right|\leqslant$$

$$\leqslant\frac{|\Phi(\alpha,\ R)|}{R}+\int_R^{\infty}\frac{|\Phi(\alpha,\ x)|}{x^2}\,dx.$$

From this and inequality (9.26) we obtain the following estimate:

$$\left|\int_R^{\infty} e^{-\alpha x}\frac{\sin x}{x}\,dx\right|\leqslant\frac{2}{R}+2\int_R^{\infty}\frac{dx}{x^2}=\frac{4}{R}. \tag{9.27}$$

This estimate implies that the integral (9.25) is uniformly convergent in α on $0\leqslant\alpha<\infty$. Indeed, let ε be an arbitrary positive number. Choose for that a number $A>0$ so that

$$\frac{4}{A}<\varepsilon.$$

It is clear that then, with $R\geqslant 1$, by the estimate (9.27) for every $\alpha\geqslant 0$

$$\left|\int_R^{\infty} e^{-\alpha x}\frac{\sin x}{x}\,dx\right|<\varepsilon$$

which implies uniform convergence in α on the half-line $0 \leqslant \alpha < \infty$ of the integral (9.25) under investigation.

2°. Use the conclusions just derived to compute the integral*

$$I = \int_0^\infty \frac{\sin x}{x}\, dx. \tag{9.28}$$

Note in the first place that the integral is the limiting value, as $\alpha \to 0+0$, of the function $I(\alpha)$ defined by relation (9.25). Indeed, the integrand in (9.25) is continuous when $\alpha \geqslant 0$ and $x \geqslant 0$ (when $x = 0$ it is assumed to be equal to unity) and the integral (9.25) uniformly converges in α on $0 \leqslant \alpha < \infty$. By Theorem 9.9 therefore the integral (9.25) is a continuous function of α on the half-line $\alpha \geqslant 0$. It follows that

$$\lim_{\alpha \to 0+0} I(\alpha) = I = \int_0^\infty \frac{\sin x}{x}\, dx. \tag{9.29}$$

We obtain for the function $I(\alpha)$ a special representation using which we shall find the value of the limit (9.29). That representation is obtained from the expression for the derivative $I'(\alpha)$. We must first therefore see if it is possible to differentiate the integral (9.25) with respect to the parameter α under the integral. To this end we check to see if the hypotheses of Theorem 9.13 hold when applied to the integral (9.25). It is obvious that the integrand and its partial derivative are continuous in α when $\alpha \geqslant 0$ and $x \geqslant 0$. We now turn to the question of uniform convergence in α of the integral

$$-\int_0^\infty e^{-\alpha x} \sin x\, dx \tag{9.30}$$

of the partial derivative of the integrand in (9.25). Fix any $\Delta > 0$. Since $|e^{-\alpha x} \sin x| \leqslant e^{-\Delta x}$ for all $\alpha \geqslant \Delta$ and since $\int_0^\infty e^{-\Delta x}\, dx$ converges, by the Weierstrass test (Theorem 9.7) the integral (9.30) converges uniformly in α when $\alpha \geqslant \Delta$. Since Δ is any positive number, we can differentiate the integral (9.25) under the integral with respect to the parameter α for any $\alpha > 0$. So when $\alpha > 0$

$$I'(\alpha) = -\int_0^\infty e^{-\alpha x} \sin x\, dx = -\frac{1}{1+\alpha^2}.$$

* The convergence of the integral under consideration was established in Section 3.1.2.

Integrating the left- and right-hand sides of the last relations we get for $\alpha > 0$

$$I(\alpha) = -\int \frac{d\alpha}{1+\alpha^2} = -\text{arc tan}\,\alpha + C. \tag{9.31}$$

We seek the constant C. Since $\left|\frac{\sin x}{x}\right| \leqslant 1$ when $x \geqslant 0$, from the expression (9.25) we get for $\alpha > 0$

$$|I(\alpha)| \leqslant \int_0^\infty e^{-\alpha x}\,dx = \frac{1}{\alpha}$$

from which it follows that

$$\lim_{\alpha\to\infty} |I(\alpha)| = 0,$$

and therefore

$$\lim_{\alpha\to\infty} I(\alpha) = 0. \tag{9.32}$$

Since $\lim\limits_{\alpha\to\infty} \text{arc tan}\,\alpha = \pi/2$, from (9.31) and (9.32) we find that $C = \pi/2$. So for $\alpha > 0$ the function $I(\alpha)$ may be represented in the following form:

$$I(\alpha) = \frac{\pi}{2} - \text{arc tan}\,\alpha.$$

From this and from formula (9.29) we evaluate the integral (9.28):

$$\int_0^\infty \frac{\sin x}{x}\,dx = \frac{\pi}{2}. \tag{9.33}$$

Remark. Consider the integral

$$K(\alpha) = \int_0^\infty \frac{\sin \alpha x}{x}\,dx. \tag{9.34}$$

We evaluate this integral for all possible α.

We make a change of variables in the integral (9.34) for $\alpha > 0$, setting $\alpha x = y$. Then

$$K(\alpha) = \int_0^\infty \frac{\sin \alpha x}{x}\,dx = \int_0^\infty \frac{\sin y}{y}\,dy = \frac{\pi}{2}.$$

We make a change of variables for $\alpha < 0$, setting $\alpha x = -y$ $(y > 0)$. Then

$$K(\alpha) = -\int_0^\infty \frac{\sin y}{y}\,dy = -\frac{\pi}{2}.$$

It is obvious that for $\alpha = 0$ the integral (9.34) is equal to zero. So

$$K(\alpha) = \int_0^\infty \frac{\sin \alpha x}{x}\,dx = \begin{cases} \pi/2 & \text{when } \alpha > 0 \\ 0 & \text{when } \alpha = 0 \\ -\pi/2 & \text{when } \alpha < 0. \end{cases}$$

The integral we have considered is commonly called *a discontinuous Dirichlet multiplier.*

Using the discontinuous Dirichlet multiplier we obtain the following analytic representation of the well-known function sgn α which is usually termed the "α sign"*:

$$\operatorname{sgn} \alpha = \frac{2}{\pi}\int_0^\infty \frac{\sin \alpha x}{x}\,dx.$$

9.4. EULER INTEGRALS

In this section we discuss some properties of important nonelementary functions called Euler integrals**.

A Euler integral of the first kind or 'beta-function' is the integral

$$\mathrm{B}(p,\ q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx. \tag{9.35}$$

In this integral p and q are assumed to be parameters. If they satisfy the conditions $p < 1$ and $q < 1$, then the integral (9.35) is an improper integral dependent on the parameters p and q, the singular points of the integral being the points $x = 0$ and $x = 1$.

A Euler integral of the second kind or 'gamma-function' is the improper integral

$$\Gamma(p) = \int_0^\infty e^{-x}x^{p-1}\,dx. \tag{9.36}$$

Note that evaluating the integral (9.36) we must take into consideration that: (1) integration is over the half-line $0 \leqslant x < \infty$, (2) when $p < 1$ the point $x = 0$ is a singular point of the integrand (the integrand goes into infinity).

In our discussion we shall take into account the above peculiarities of the functions $B(p,\ q)$ and $\Gamma(p)$. We shall see below that the integrals (9.35) and (9.36) converge for $p > 0$ and $q > 0$.

---

* This term is due to the fact that the values of sgn α for $\alpha > 0$, $\alpha = 0$, and $\alpha < 0$ equal 1, 0, -1 respectively.

** For a detailed account of Euler integrals see: E.G. Whittaker and G.N. Watson, "A Course of Modern Analysis". An introduction to the general theory of infinite processes and analytic functions; with an account of the principal transcendental functions. 4th ed., Cambridge, 1927.

9.4.1. The domain of convergence of Euler integrals. We prove that the function $B(p, q)$ is defined for all positive values of the parameters p and q and that $\Gamma(p)$ is defined for all positive values of p.

We first consider $B(p, q)$. For $p \geqslant 1$ and $q \geqslant 1$ the integrand in relation (9.35) is continuous and the integral on the right of (9.35) is therefore proper. Thus $B(p, q)$ is defined for all values of p and q mentioned. We now turn to the case where one or both of the following inequalities hold:

$$0 < p < 1, \; 0 < q < 1. \tag{9.37}$$

In this case one or both of the points $x = 0$ and $x = 1$ are singular points of the integrand. With this in mind we represent $B(p, q)$ as follows:

$$B(p, q) = \int_0^{1/2} x^{p-1}(1-x)^{q-1}\,dx + \int_{1/2}^{1} x^{p-1}(1-x)^{q-1}\,dx =$$
$$= I_1(p, q) + I_2(p, q).$$

Obviously, either of the integrals $I_1(p, q)$ and $I_2(p, q)$ has only one singular point.

For the integral $I_1(p, q) = \int_0^{1/2} x^{p-1}(1-x)^{q-1}\,dx$ the singular point is the point $x = 0$. Noting that on the closed interval $[0, 1/2]$ the function $(1-x)^{q-1}$ is continuous and is therefore bounded by some constant C, it is easy to see that the function Cx^{p-1} is majorant for the integrand of $I_1(p, q)$. It follows that $I_1(p, q)$ converges for $0 < p < 1$ and any q. Reasoning similarly it is easy to see that $I_2(p, q)$ converges for $0 < q < 1$ and any p.

So we have seen that in the case where the inequalities $p > 0$ and $q > 0$ hold the integral (9.35) converges, i.e. *the function* $B(p, q)$ *is defined for all positive values of* p *and* q.

We now turn to the function $\Gamma(p)$. We have already noted that the integral (9.36) has two types of peculiarities: integration over a half-line and a singular point $x = 0$. To separate these peculiarities we divide the domain of integration into two parts so that there is only one of those peculiarities in either. For example, we may represent $\Gamma(p)$ as follows:

$$\Gamma(p) = \int_0^1 e^{-x}x^{p-1}\,dx + \int_1^{\infty} e^{-x}x^{p-1}\,dx = I_1(p) + I_2(p).$$

Since $|e^{-x}x^{p-1}| \leqslant x^{p-1}$ for $x > 0$, by the partial comparison test $I_1(p)$ converges when $p > 0$. The integral $I_2(p)$ also converges

when $p > 0$. To show this we can use the partial comparison test in the limit form: $\lim\limits_{x\to\infty} e^{-x}x^r = 0$ for any r. So we have proved that the domain of $\Gamma(p)$ is the half-line $p > 0$.

9.4.2. The continuity of Euler integrals. We prove that $\mathrm{B}(p, q)$ is continuous in the quadrant $p > 0$, $q > 0$ and that $\Gamma(p)$ is continuous on the half-line $p > 0$. We first consider $\mathrm{B}(p, q)$. To prove that it is continuous in the quadrant $p > 0$, $q > 0$ it is obviously sufficient to show that the integral (9.35) is uniformly convergent with respect to the parameters p and q when $p \geqslant p_0 > 0$ and $q \geqslant q_0 > 0$ for any fixed positive values of p_0 and q_0. Since $p_0 - 1 \leqslant p - 1$, $q_0 - 1 \leqslant q - 1$, for $0 < x < 1$

$$x^{p-1}(1-x)^{q-1} \leqslant x^{p_0-1}(1-x)^{q_0-1}.$$

From these inequalities and from the convergence of $\int\limits_0^1 x^{p_0-1}(1-x)^{q_0-1}\,dx$ it follows by virtue of the Weierstrass test that the integral (9.35) is uniformly convergent for the indicated values of p and q. Thus the continuity of $\mathrm{B}(p, q)$ for $p > 0$ and $q > 0$ is proved.

To prove that $\Gamma(p)$ is continuous on the half-line $p > 0$ it is obviously sufficient to establish that the integral (9.36) is uniformly convergent with respect to p when $0 < p_0 \leqslant p \leqslant p_1$ for any fixed values of p_0 and p_1 satisfying the condition $0 < p_0 < p_1$. Since for the indicated values of p, p_0, and p_1 and for $x > 0$

$$e^{-x}x^{p-1} \leqslant e^{-x}\,[x^{p_0-1} + x^{p_1-1}],$$

the convergence of the integral

$$\int\limits_0^\infty e^{-x}\,[x^{p_0-1} + x^{p_1-1}]\,dx$$

implies, by the Weierstrass test, the uniform convergence of the integral (9.36) for the indicated values of p. Thus the continuity of $\Gamma(p)$ for $p > 0$ is proved.

9.4.3. Some properties of the function $\Gamma(p)$. Here we shall prove that $\Gamma(p)$ has a derivative of any order. We shall also obtain for $\Gamma(p)$ a formula called *a reduction formula*.

Differentiating $\Gamma(p)$ with respect to the parameter under the integral we obtain the following integral

$$\int\limits_0^\infty x^{p-1}e^{-x}\ln x\,dx \tag{9.38}$$

which converges uniformly in p on any closed interval $0 < p_0 \leqslant p \leqslant p_1$. Indeed, the absolute value of the integrand in the integral

(9.38) satisfies on $0 < x < \infty$ the inequality

$$|x^{p-1}e^{-x}\ln x| < e^{-x}\,|\ln x|\,(x^{p_0-1}+x^{p_1-1}).$$

Hence the convergence of the integral

$$\int_0^\infty e^{-x}\,|\ln x|\,(x^{p_0-1}+x^{p_1-1})\,dx$$

implies according to the Weierstrass test the uniform convergence of the integral (9.38). This fact, together with the continuity of the integrand in the integral (9.38)* when $0 < x < \infty$, $0 < p < \infty$, allows us to conclude that it is possible to differentiate $\Gamma(p)$ with respect to the parameter under the integral. So *the derivative* $\Gamma'(p)$ *exists and is equal to the expression* (9.38).

Reasoning similarly it is easy to show that $\Gamma(p)$ has a derivative of any order and that that derivative can be found by differentiating with respect to the parameter p under the integral in the expression (9.36) for $\Gamma(p)$.

We now proceed to derive a *reduction formula* for $\Gamma(p)$.

Applying the formula of integration by parts to $\Gamma(p+1)$, with $p > 0$, we get

$$\Gamma(p+1)=\int_0^\infty x^p e^{-x}\,dx=[-x^p e^{-x}]_0^\infty+p\int_0^\infty x^{p-1}e^{-x}\,dx=p\Gamma(p).$$

So for any $p > 0$

$$\Gamma(p+1)=p\Gamma(p). \tag{9.39}$$

Applying successively formula (9.39) for any $p > n-1$ and any natural n we get

$$\Gamma(p+1)=p(p-1)\ldots(p-n+1)\,\Gamma(p-n+1). \tag{9.40}$$

Relation (9.40) is called *a reduction formula* for $\Gamma(p)$. Using (9.40) the gamma-function for values of the independent variable greater than unity is "reduced" to the gamma-function for values of the independent variable between zero and unity.

Since $\Gamma(1)=\int_0^\infty e^{-x}\,dx=1$, setting in (9.40) $p=n$ we get

$$\Gamma(n+1)=n(n-1)\ldots 2\cdot 1=n!$$

This formula will be used below in deriving the so-called Stirling** formula giving an asymptotic representation for $n!$

The information obtained for the function $\Gamma(p)$ allows us to characterize qualitatively the graph of this function. We shall

* This function is a partial derivative with respect to p of the integrand in the expression (9.36) for $\Gamma(p)$.

** James Stirling (1692-1770) is a Scottish mathematician.

make a geometrical study of the graph of $\Gamma(p)$ following in the main the pattern presented in Section 9.6 of [1].

We have established that the domain of $\Gamma(p)$ is the half-line $0 < p < \infty$. The function $\Gamma(p)$ is continuous and differentiable any number of times on the half-line, it being possible to find any derivative by differentiating the expression (9.36) for $\Gamma(p)$ with respect to the parameter p under the integral. In particular, the second derivative $\Gamma''(p)$ is

$$\Gamma''(p) = \int_0^\infty x^{p-1} (\ln x)^2 e^{-x}\, dx.$$

Since $\Gamma''(p) > 0$, the first derivative $\Gamma'(p)$ may have only one zero. Since $\Gamma(1) = \Gamma(2)$*, by the Rolle theorem that zero of $\Gamma'(p)$ exists and is in the interval $(1, 2)$. Since $\Gamma''(p) > 0$, at the point where $\Gamma'(p)$ vanishes $\Gamma(p)$ has a minimum. Note also that the graph of $\Gamma(p)$ is convex down. The graph of $\Gamma(p)$ has a vertical asymptote at the point $p = 0$. Indeed, since $\Gamma(1) = 1$ and $\Gamma(p) = \Gamma(p+1)/p$, the continuity of $\Gamma(p)$ at the point 1 implies that $\Gamma(p) \to +\infty$ as $p \to 0+0$. Obviously, $\Gamma(p) \to +\infty$ as $p \to +\infty$. Note without proof that the graph of $\Gamma(p)$ has no inclined asymptotes.

9.4.4. Some properties of the function $B(p, q)$. Here we shall establish the symmetry properties of $B(p, q)$ and its reduction formula.

We make a change of variable in the integral (9.35), setting $x = 1 - t$. On making the necessary computations we shall see that

$$B(p, q) = B(q, p) \tag{9.41}$$

which expresses *the symmetry property of the function* $B(p, q)$.

We establish reduction formulas for $B(p, q)$. To do this we turn to the function $B(p, q+1)$, assuming p and q to be positive. Applying integration by parts and the formula $x^p = x^{p-1} - x^{p-1}(1-x)$ we get

$$B(p, q+1) = \int_0^1 x^{p-1} (1-x)^q\, dx =$$

$$= \left[\frac{x^p}{p} (1-x)^q\right]_0^1 + \frac{q}{p} \int_0^1 x^p (1-x)^{q-1}\, dx =$$

$$= \frac{q}{p} \int_0^1 \{x^{p-1} (1-x)^{q-1} - x^{p-1} (1-x)^q\}\, dx =$$

$$= \frac{q}{p} B(p, q) - \frac{q}{p} B(p, q+1).$$

* This follows from relation (9.39).

From these relations we obtain the following formula:

$$B(p, q+1) = \frac{q}{p+q} B(p, q). \tag{9.42}$$

Quite similarly for $p > 0$ and $q > 0$

$$B(p+1, q) = \frac{p}{p+q} B(p, q). \tag{9.43}$$

Formulas (9.42) and (9.43) are called reduction formulas for the function $B(p, q)$. Successive application of these formulas reduces computing $B(p, q)$ for arbitrary positive values of the independent variables to computing $B(p, q)$ for values of the independent variables in the half-open square $0 < p \leqslant 1$, $0 < q \leqslant 1$.

9.4.5. The relation between Euler integrals. We make a change of variable in the integral (9.35), setting $x = 1/(1+t)$. As a result we obtain for $B(p, q)$ the following expression:

$$B(p, q) = \int_0^\infty \frac{t^{q-1}}{(1+t)^{p+q}} dt. \tag{9.44}$$

Using formula (9.41), along with (9.44) we obtain the following expression for $B(p, q)$:

$$B(p, q) = \int_0^\infty \frac{t^{p-1}}{(1+t)^{p+q}} dt. \tag{9.45}$$

We now turn to the expression (9.36) for $\Gamma(p)$. Using the substitution $x = ty$, $t > 0$, we convert that expression to the form

$$\frac{\Gamma(p)}{t^p} = \int_0^\infty e^{-ty} y^{p-1} dy. \tag{9.46}$$

Replacing in this formula p by $p+q$ and t by $1+t$ we get

$$\frac{\Gamma(p+q)}{(1+t)^{p+q}} = \int_0^\infty e^{-(1+t)y} y^{p+q-1} dy.$$

We multiply both sides of the last equation by t^{p-1} and integrate with respect to t between 0 and ∞. Obviously, according to relation (9.45) we obtain the formula

$$\Gamma(p+q) B(p, q) = \int_0^\infty dt \int_0^\infty y^{p+q-1} t^{p-1} e^{-(1+t)y} dy. \tag{9.47}$$

If on the right of relation (9.47) we may interchange the order of integration with respect to t and y, then, taking into account (9.46),

we get

$$\Gamma(p+q)\,\mathrm{B}(p,\ q)=\int_0^\infty y^{p+q-1}e^{-y}\,dy\int_0^\infty t^{p-1}e^{-ty}\,dt=$$

$$=\int_0^\infty y^{p+q-1}e^{-y}\,\frac{\Gamma(p)}{y^p}\,dy=\Gamma(p)\int_0^\infty y^{q-1}e^{-y}\,dy=\Gamma(p)\,\Gamma(q),$$

i.e. prove the validity of the formula

$$\mathrm{B}(p,\ q)=\frac{\Gamma(p)\cdot\Gamma(q)}{\Gamma(p+q)}. \tag{9.48}$$

We shall now show that it is possible to change the order of integration on the right of (9.47). To do this it is necessary to verify that the hypotheses of Theorem 9.12 hold. First let $p>1$ and $q>1$. Then obviously the hypotheses of Theorem 9.12 hold. Indeed,

(1) the function $f(t,\ y)=t^{p-1}y^{p+q-1}e^{-(1+t)y}$ is nonnegative and continuous in the quadrant $t\geqslant 0,\ y\geqslant 0$.

(2) The integral $\int_0^\infty f(t,\ y)\ dy = t^{p-1}\int_0^\infty y^{p+q-1}e^{-(1+t)y}\,dy = \frac{\Gamma(p+q)\,t^{p-1}}{(1+t)^{p+q}}$ is a continuous function of t for $t\geqslant 0$.

(3) The integral $\int_0^\infty f(t,\ y)\,dt=y^{p+q-1}e^{-y}\int_0^\infty t^{p-1}e^{-ty}\,dy=\Gamma(p)y^{q-1}e^{-y}$ is a continuous function of y for $y\geqslant 0$.

(4) The convergence of $\int_0^\infty dy\int_0^\infty f(t,\ y)\,dt$ is established by straightforward computation.

So for $p>1$ and $q>1$ formula (9.48) is true. But if only the conditions $p>0$ and $q>0$ hold, then by that which has been proved

$$\mathrm{B}(p+1,\ q+1)=\frac{\Gamma(p+1)\,\Gamma(p+1)}{\Gamma(p+q+2)}.$$

Using the reduction formulas for B $(p,\ q)$ and Γ (p) we again obtain from this (9.48).

9.4.6. Evaluation of definite integrals by means of Euler integrals. Euler integrals are well-studied nonelementary functions. A problem is considered to be solved if it can be reduced to computing Euler integrals.

Here are some examples of computing ordinary and improper integrals by reducing them to Euler integrals.

1. Compute the integral

$$I=\int_0^{\infty} x^{1/4}(1+x)^{-2}\,dx.$$

Turning to formulas (9.44) and (9.48) we obviously get

$$I=\mathrm{B}\left(\frac{5}{4},\ \frac{3}{4}\right)=\frac{\Gamma(5/4)\,\Gamma(3/4)}{\Gamma(2)}=\frac{1}{4}\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{3}{4}\right).$$

2. Compute the integral

$$I=\int_0^{\pi/2}\sin^{p-1}\varphi\cos^{q-1}\varphi\,d\varphi.$$

Setting $x=\sin^2\varphi$ we get

$$I=\frac{1}{2}\int_0^1 x^{\frac{p}{2}-1}(1-x)^{\frac{q}{2}-1}\,dx=$$

$$=\frac{1}{2}\mathrm{B}\left(\frac{p}{2},\ \frac{q}{2}\right)=\frac{1}{2}\,\frac{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}{\Gamma\left(\frac{p+q}{2}\right)}.$$

3. Consider the integral

$$I_{p-1}=\int_0^{\pi/2}\sin^{p-1}\varphi\,d\varphi.$$

Using the result obtained in Example 2 (it is necessary to set $q=1$) we find

$$\int_0^{\pi/2}\sin^{p-1}\varphi\,d\varphi=\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)}. \tag{9.49}$$

We further have

$$\Gamma\left(\frac{1}{2}\right)=\int_0^{\infty}e^{-x}\frac{dx}{\sqrt{x}}=2\int_0^{\infty}e^{-(\sqrt{x})^2}\,d\sqrt{x}.$$

Setting $\sqrt{x}=t$ and noticing that $\int_0^{\infty}e^{-t^2}\,dt$ equals $1/2\int_{-\infty}^{\infty}e^{-t^2}\,dt$ we get, according to the example considered in Section 3.4.2 (Poisson integral),

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}.$$

Formula (9.49) therefore becomes

$$I_{p-1}=\int_0^{\pi/2}\sin^{p-1}\varphi\,d\varphi=\frac{\sqrt{\pi}}{2}\frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)}. \tag{9.50}$$

9.5. THE STIRLING FORMULA

The term "Stirling formula" is applied to the following asymptotic formula:

$$n!=\sqrt{2\pi n}\,n^n e^{-n}(1+\alpha_n), \tag{9.51}$$

where $\alpha_n\to 0$ as $n\to\infty$.

We shall prove here a more general formula describing as precisely as desired the behaviour of the Euler gamma-function for large values of the independent variable:

$$\Gamma(\lambda+1)=\int_0^{\infty}t^{\lambda}e^{-t}\,dt. \tag{9.52}$$

To do this we use the so-called *Laplace method* based on the following statement.

Lemma. Let $f(t)$ be a function integrable for some $a>0$ on a closed interval $[-a, a]$ and representable as

$$f(t)=\sum_{k=0}^{2n-1}c_k t^k+O(t^{2n}). \tag{9.53}$$

Then the following asymptotic formula holds:

$$\int_{-a}^{a}e^{-\lambda t^2}f(t)\,dt=\sum_{m=0}^{n-1}c_{2m}\frac{\Gamma\left(m+\frac{1}{2}\right)}{\lambda^{m+\frac{1}{2}}}+\frac{O(1)}{\lambda^{n+\frac{1}{2}}}. \tag{9.54}$$

Proof. Substitute relation (9.53) in the integral on the left of (9.54) and take into account the fact that the integrals corresponding to odd powers of t vanish. To evaluate the remaining integrals it suffices to show that for $m\geqslant 0$

$$\int_0^{a}t^{2m}e^{-\lambda t^2}\,dt=\frac{1}{2}\frac{\Gamma\left(m+\frac{1}{2}\right)}{\lambda^{m+\frac{1}{2}}}+O(e^{-\lambda a^2}). \tag{9.55}$$

Represent the integral on the left of (9.55) in the following form:

$$\int_0^a t^{2m}e^{-\lambda t^2}\,dt=\int_0^\infty t^{2m}e^{-\lambda t^2}\,dt-\int_a^\infty t^{2m}e^{-\lambda t^2}\,dt. \tag{9.56}$$

Make the substitution $x = \lambda t^2$ in the first integral on the right of (9.56) to get

$$\int_0^\infty t^{2m}e^{-\lambda t^2}\,dt\ \frac{1}{2\lambda^{m+\frac{1}{2}}}\int_0^\infty x^{m-\frac{1}{2}}e^{-x}\,dx=\frac{1}{2}\,\frac{\Gamma\left(m+\frac{1}{2}\right)}{\lambda^{m+\frac{1}{2}}}. \tag{9.57}$$

Further note that for $\lambda > 1$ and $t \geqslant a$

$e^{-\lambda t^2}\leqslant e^{-(\lambda-1)a^2}e^{-t^2}.$

Applying this inequality, evaluate the second integral on the right of (9.56)

$$\int_a^\infty e^{-\lambda t^2}t^{2m}\,dt\leqslant e^{-(\lambda-1)a^2}\int_a^\infty t^{2m}e^{-t^2}\,dt=ce^{-\lambda a^2}. \tag{9.58}$$

Equations (9.56), (9.57) and the estimate (9.58) yield the required formula (9.55), which proves the lemma.

To apply the lemma make the substitution $t = \lambda(1 + x)$ in the integral (9.52). As a result this becomes

$$\Gamma(\lambda+1)=\lambda^{\lambda+1}e^{-\lambda}\int_{-1}^\infty e^{-\lambda[x-\ln(1+x)]}\,dx. \tag{9.59}$$

Denote by $g(x)$ the following function defined on the half-line $x > -1$:

$$g(x)=\operatorname{sgn} x\cdot\sqrt{x-\ln(1+x)}. \tag{9.60}$$

Then equation (9.59) may be rewritten in the form

$$\Gamma(\lambda+1)=\lambda^{\lambda+1}e^{-\lambda}\int_{-1}^\infty e^{-\lambda g^2(x)}\,dx. \tag{9.61}$$

Our aim is to study the asymptotic behaviour of the following integral as $\lambda\to+\infty$:

$$I(\lambda)=\int_{-1}^\infty e^{-\lambda g^2(x)}\,dx. \tag{9.62}$$

To do this we consider in more detail the function $g(x)$ defined by equation (9.60). Since

$$\frac{d}{dx} g^2(x) = \frac{d}{dx}(x - \ln(1+x)) = \frac{x}{1+x}, \tag{9.63}$$

the function $g^2(x)$ is strictly decreasing when $-1 < x < 0$ and strictly increasing when $x > 0$. It follows that $g(x)$ is strictly increasing on the half-line $x > -1$, its range of values being the entire number line. Further, since $g^2(x)$ has in the neighbourhood of the point $x = 0$ the expansion

$$g^2(x) = x - \ln(1+x) = x - \left(x - \frac{x^2}{2} + O(x^3)\right) = \frac{x^2}{2} + O(x^3),$$

there is a function $h(x)$ strictly positive for $x > -1$ such that

$g^2(x) = x^2 h(x)$.

The function $h(x)$ is infinitely differentiable for $x > -1$, consequently so is the function $g(x) = x\sqrt{h(x)}$.

Considering the foregoing one may say that for the function $y = g(x)$ defined by equation (9.60) there is an inverse function $x = g^{-1}(y)$ strictly increasing and infinitely differentiable on the entire number line and satisfying the condition $g^{-1}(0) = 0$.

Denote that inverse function by $x = \varphi(y)$. Using the above properties of the function find the asymptotics of the integral (9.62). The following theorem is true.

Theorem 9.13. Let $x = \varphi(y)$ *be the inverse of the function* $y = g(x)$ *defined by equation* (9.60). *Then for the integral* (9.62) *the following asymptotic formula is true with any* n:

$$I(\lambda) = \sum_{m=0}^{n-1} \frac{\varphi^{(2m+1)}(0)}{(2m)!} \frac{\Gamma\left(m+\frac{1}{2}\right)}{\lambda^{m+\frac{1}{2}}} + \frac{O(1)}{\lambda^{n+\frac{1}{2}}}. \tag{9.64}$$

Proof. Fix an arbitrary positive number a and set $b = \varphi(-a)$, $c = \varphi(a)$. This means that $a = g(c) = -g(b)$ and therefore $-1 < b < 0$ and $c > 0$.

Evaluate the following two integrals

$$I_1(\lambda) = \int_{-1}^{b} e^{-\lambda g^2(x)}\, dx, \quad I_2(\lambda) = \int_{c}^{\infty} e^{-\lambda g^2(x)}\, dx. \tag{9.65}$$

To evaluate the first integral note that for $-1 < x < b$ we have $g(x) < -a$, i.e. $g^2(x) > a^2$, and therefore

$$e^{-\lambda g^2(x)} < e^{-\lambda a^2}.$$

In such a case

$$I_1(\lambda) \leqslant e^{-\lambda a^2} \int_{-1}^{b} dx = (1 - |b|) e^{-\lambda a^2}. \tag{9.66}$$

The evaluation of $I_2(\lambda)$ is similar. For $x > c$ we have $g(x) > a$, i.e. $g^2(x) > a^2$. Therefore, for $\lambda > 1$ and $x > c$

$$e^{-\lambda g^2(x)} = e^{-(\lambda-1)g^2(x)} e^{-g^2(x)} < e^{-(\lambda-1)a^2} e^{-g^2(x)}.$$

From this we get

$$I_2(\lambda) \leqslant e^{-(\lambda-1)a^2} \int_{c}^{\infty} e^{-g_2(x)}\, dx = c_1 e^{-\lambda a^2}. \tag{9.67}$$

From the estimates (9.66) and (9.67) satisfied by the integrals (9.65) we obtain for the integral (9.62) the following relation:

$$I(\lambda) = \int_{b}^{c} e^{-\lambda g^2(x)}\, dx + O(e^{-\lambda a^2}). \tag{9.68}$$

Make a change of variable $t = g(x)$ in the integral (9.68), i.e. $x = \varphi(t)$. As a result we get

$$I(\lambda) = \int_{-a}^{a} e^{-\lambda t^2} \varphi'(t)\, dt + O(e^{-\lambda a^2}). \tag{9.69}$$

Since the function $\varphi'(t)$ is infinitely differentiable, use the Maclaurin formula to represent $\varphi'(t)$ as

$$\varphi'(t) = \sum_{k=0}^{2n-1} \frac{\varphi^{(k+1)}(0)}{k!} t^k + O(t^{2n}).$$

To obtain formula (9.64) it remains to apply the lemma to the function $f(t) = \varphi'(t)$. Thus Theorem 9.13 is proved.

In conclusion we describe the following simple way of computing $\varphi^{(k)}(0)$. From (9.63) we get

$$2g \cdot g' = \frac{x}{x+1} = \frac{\varphi(t)}{\varphi(t)+1}.$$

This implies the relation

$$\varphi'(t) = \frac{1}{g'} = 2g\, \frac{1+\varphi(t)}{\varphi(t)} = 2t\, \frac{1+\varphi(t)}{\varphi(t)}.$$

We thus obtain the following equation:

$$\varphi(t)\, \varphi'(t) = 2t + 2t\varphi(t). \tag{9.70}$$

Differentiating successively (9.70) and setting $t = 0$ we determine all $\varphi^{(k)}(0)$. As an illustration, we evaluate the first three derivatives of $\varphi(t)$ at zero.

On differentiating (9.70) we get

$$[\varphi'(t)]^2 + \varphi(t)\varphi''(t) = 2 + 2(t\varphi'(t) + \varphi(t)). \quad (9.71)$$

Set $t = 0$ and consider that $\varphi(0) = 0$. Then $\varphi'^2(0) = 2$, i.e. $\varphi'(0) = \sqrt{2}$.

After differentiating equation (9.71) we get

$$3\varphi'\cdot\varphi'' + \varphi\cdot\varphi''' = 2(t\varphi'' + 2\varphi').$$

On equating t to zero we get $3\sqrt{2}\,\varphi''(0) = 4\sqrt{2}$, i.e. $\varphi''(0) = 4/3$. Similarly from

$$3\varphi''^2 + 4\varphi'\cdot\varphi''' + \varphi\cdot\varphi^{\mathrm{IV}} = 2(t\varphi''' + 3\varphi'')$$

we get $\varphi'''(0) = \sqrt{2}/3$.

Consequently formula (9.64) may be written as

$$I(\lambda) = \sqrt{2}\,\frac{\Gamma(1/2)}{\sqrt{\lambda}} + \frac{\sqrt{2}}{6}\,\frac{\Gamma(3/2)}{\lambda\sqrt{\lambda}} + \frac{O(1)}{\lambda^2\sqrt{\lambda}}. \quad (9.72)$$

Substitute equation (9.72) in (9.61) and consider that $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(3/2) = 1/2\Gamma(1/2) = \sqrt{\pi}/2$. As a result we get

$$\Gamma(\lambda+1) = \sqrt{2\pi\lambda}\,\lambda^{\lambda}e^{-\lambda}\left(1 + \frac{1}{12\lambda} + \frac{O(1)}{\lambda^2}\right). \quad (9.73)$$

Write out the first five terms of the asymptotic expansion of the Euler gamma-function:

$$\Gamma(\lambda+1) = \sqrt{2\pi\lambda}\,\lambda^{\lambda}e^{-\lambda}\left(1 + \frac{1}{12\lambda} + \frac{1}{288\lambda^2} - \right.$$
$$\left. - \frac{139}{51\,840\lambda^3} - \frac{571}{2\,488\,320\lambda^4} + \frac{O(1)}{\lambda^5}\right).$$

Note without proof that the remainder of an asymptotic series does not exceed the last term retained.

9.6. PARAMETER-DEPENDENT MULTIPLE INTEGRALS

9.6.1. Parameter-dependent proper multiple integrals. Let $x = (x_1, x_2, \ldots, x_m)$ be an arbitrary point in a domain D of an m-dimensional Euclidean space E^m and let $y = (y_1, y_2, \ldots, y_l)$ be a point in a domain Ω of a space E^l. Denote by $D \times \Omega$ a subset of an $(l + m)$-dimensional Euclidean space consisting of all points $z = (z_1, z_2, \ldots, z_{m+l})$ such that the point $(z_1, z_2, \ldots, z_m)$ is in D and the point $(z_{m+1}, z_{m+2}, \ldots, z_{m+l})$ is in Ω. We shall often use the notation $z = (x, y) \in D \times \Omega$. The closure of a domain D

will be designated $\bar{D}$. It is easy to see that the closure of $D \times \Omega$ coincides with $\bar{D} \times \bar{\Omega}$.

Let $f(x, y)$ be a function defined in $D \times \Omega$, the function $f(x, y_0)$ being integrable over x for any $y_0 \in \Omega$ in D. Then the function

$$I(y) = \int_D f(x, y)\, dx \tag{9.74}$$

defined in Ω will be called the integral dependent on the parameter y. Notice that the parameter y is an l-dimensional vector and the integral (9.74) is therefore dependent on l numerical parameters $y_1, y_2, \ldots, y_l$.

In close analogy with Theorems 9.9 to 9.12 the following theorems are proved.

Theorem 9.14 (on continuity of the integral (9.74) in the parameter). If $f(x, y)$ is continuous in a collection of independent variables in a closed domain $\bar{D} \times \bar{\Omega}$, then the integral (9.74) *is a continuous function of the parameter y in the domain $\bar{\Omega}$.*

Theorem 9.15 (on integration of the integral (9.74) with respect to the parameter). If $f(x, y)$ is continuous in a collection of independent variables in a closed domain $\bar{D} \times \bar{\Omega}$, then the function (9.74) *can be integrated with respect to the parameter under the integral, i.e.*

$$\int_\Omega I(y)\, dy = \int_D dx \int_\Omega f(x, y)\, dy.$$

Theorem 9.16 (on differentiability of the integral (9.74) with respect to the parameter). If $f(x, y)$ and its partial derivative $\frac{\partial f}{\partial y_k}$ are continuous in $\bar{D} \times \bar{\Omega}$, then the integral (9.74) *has a continuous partial derivative $\frac{\partial I}{\partial y_k}$ in Ω, with*

$$\frac{\partial I}{\partial y_k} = \int_D \frac{\partial f(x, y)}{\partial y_k}\, dx.$$

9.6.2. Parameter-dependent improper multiple integrals. The concept of parameter-dependent improper multiple integral could be introduced, as in the preceding subsection, for the case where the integrand $f(x, y)$ is defined in $D \times \Omega$, where $D \subset E^m$ and $\Omega \subset E^l$. Of the greatest interest, however, is the case $D = \Omega$ to be studied here. We shall also assume that $f(x, y) = F(x, y)\, g(x)$, where $F(x, y)$ is continuous for $x \neq y$ in $\bar{D} \times \bar{D}$ and $g(x)$ is bounded in D. Thus we are considering integrals of the form

$$V(y) = \int_D F(x, y)\, g(x)\, dx \tag{9.75}$$

where the integrand may have singularities for $x = y$ only. We shall be concerned with continuity of integrals of the form (9.75) in the parameter y. In this connection we introduce the following definition of *uniform convergence of the integral* (9.75) *at a point*. We shall use $K(y_0, \delta)$ to denote a ball of radius δ with centre at a point y_0.

Definition. An integral (9.75) *is said to be* convergent uniformly in the parameter y at a point $y_0 \in D$ *if given any* $\varepsilon > 0$ *we can find* $\delta > 0$ *such that* $K(y_0, \delta) \subset D$ *and for any cubable domain* $\omega \subset \subset K(y_0, \delta)$ *and each point* $y \in K(y_0, \delta)$

$$\left| \int_\omega F(x, y) g(x)\, dx \right| < \varepsilon.$$

Theorem 9.17. If the integral (9.75) *converges uniformly in y at a point* $y_0 \in D$, *then it is continuous at the point* y_0.

Proof. We want to prove that for any $\varepsilon > 0$ there is $\delta > 0$ such that $|V(y) - V(y_0)| < \varepsilon$ when $|y - y_0| < \delta$. From the definition of uniform convergence at a point it follows that there is $\delta_1 > 0$ such that $K(y_0, \delta_1) \subset D$ and that for $y \in K(y_0, \delta_1)$

$$\left| \int_K F(x, y) g(x)\, dx \right| < \varepsilon/3. \tag{9.76}$$

Set

$$\begin{aligned} V_1(y) &= \int_K F(x, y) g(x)\, dx, \\ V_2(y) &= \int_{D/K} F(x, y) g(x)\, dx. \end{aligned} \tag{9.77}$$

From inequality (9.76) it follows that for $|y - y_0| < \delta_1$

$$|V_1(y)| < \varepsilon/3. \tag{9.78}$$

Further note that for $x \in D \setminus K(y_0, \delta_1)$ and $y \in K(y_0, \delta_1/2)$ the function $F(x, y)$ is uniformly continuous in a collection of independent variables. Hence there is a positive number $\delta < \delta_1/2$ such that for $|y - y_0| < \delta$

$$|F(x, y_0) - F(x, y)| < \varepsilon/3M\,|D|,$$

where M is a constant bounding the function g and $|D|$ is the volume of a domain D. In such a case for $|y - y_0| < \delta$

$$|V_2(y) - V_2(y_0)| \leqslant M \int_{D\setminus K(y_0, \delta_1)} |F(x, y_0) - F(x, y)|\, dx \leqslant \varepsilon/3. \tag{9.79}$$

From relations (9.77) to (9.79) it follows that for $|y - y_0| < \delta$
$$|V(y) - V(y_0)| \leqslant |V_1(y)| + |V_1(y_0)| + |V_2(y) - V_2(y_0)| < \varepsilon.$$
Thus the theorem is proved.

We show one sufficient condition for the uniform convergence of an integral at a point that is most commonly occurring in applications.

Theorem 9.18. Let $F(x, y)$ be a function continuous in $\bar{D} \times \bar{D}$ for $x \neq y$ and let $g(x)$ be a function uniformly bounded in D. Suppose that there are constants λ, $0 < \lambda < m$, and $c > 0$ such that for all $x \in D$, $y \in D$

$$|F(x, y)| \leqslant c\,|x - y|^{-\lambda}. \tag{9.80}$$

Then the integral (9.75) *converges uniformly in y at each point $y_0 \in D$.*

Proof. Let y_0 be an arbitrary point of D. We want to prove that given any $\varepsilon > 0$ we can find $\delta > 0$ such that for any cubable domain $\omega \subset K(y_0, \delta)$ and each $y \in K(y_0, \delta)$

$$\left|\int_\omega F(x, y)\, g(x)\, dx\right| < \varepsilon. \tag{9.81}$$

Using (9.80) and the condition for the boundedness of $g(x)$ we get

$$\left|\int_\omega F(x, y)\, g(x)\, dx\right| \leqslant c_1 \int_\omega |x - y|^{-\lambda}\, dx.$$

Fix a point $y \in K(y_0, \delta)$ and note that the condition $\omega \subset K(y_0, \delta)$ implies the inclusion $\omega \subset K(y, 2\delta)$. Hence

$$\left|\int_\omega F(x, y)\, g(x)\, dx\right| \leqslant c_1 \int_{K(y,\, 2\delta)} |x - y|^{-\lambda}\, dx. \tag{9.82}$$

Transforming in the integral on the right of (9.82) to spherical coordinates with centre at a point y (see Section 2.5.3°) we get

$$\left|\int_\omega F(x, y)\, g(x)\, dx\right| \leqslant c_2 \int_0^{2\delta} r^{m-1-\lambda}\, dr = \frac{c_2 2^{m-\lambda}}{m - \lambda}\, \delta^{m-\lambda} = c_3 \delta^{m-\lambda}.$$

It follows that on choosing δ to be sufficiently small we obtain inequality (9.81). Thus the theorem is proved.

9.6.3. Application to Newton potential theory. Suppose a mass m_0 is placed at some point $P_0(x, y, z)$. By the law of gravitation a mass m placed at a point $M(\xi, \eta, \zeta)$ is acted on by a force

$$F = -\gamma \frac{m m_0}{R^2} \bar{r},$$

where $R = \rho(P_0, M) = \sqrt{(x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2}$, γ is the gravitational constant, and $\bar{r} = \bar{R}/R$ is a unit vector whose

direction coincides with that of $\overline{P_0M}$. Assuming $\gamma = 1$ and the mass $m = 1$ we obtain the gravitational force

$$F = -\frac{m_0}{R}\bar{r}.$$

Note that the components of this force are of the form

$$X = -\frac{m_0}{R^3}(\xi - x),$$

$$Y = -\frac{m_0}{R^3}(\eta - y),$$

$$Z = -\frac{m_0}{R^3}(\zeta - z).$$

It is obvious that the potential of gravitation defined as a scalar function u such that $F = \operatorname{grad} u$ equals

$$u = \frac{m_0}{R}.$$

If the mass is concentrated not at a point $P_0(x, y, z)$ but is distributed over a domain D with density $\rho(x, y, z)$, then for the potential of gravitation and for the components of gravitation we obtain the following expression:

$$u(\xi, \eta, \zeta) = \iiint\limits_D \frac{\rho(x, y, z)}{R}\,dx\,dy\,dz, \tag{9.83}$$

$$\left.\begin{aligned} X &= -\iiint\limits_D \frac{\rho(x, y, z)}{R^3}(\xi - x)\,dx\,dy\,dz, \\ Y &= -\iiint\limits_D \frac{\rho(x, y, z)}{R^3}(\eta - y)\,dx\,dy\,dz, \\ Z &= -\iiint\limits_D \frac{\rho(x, y, z)}{R^3}(\zeta - z)\,dx\,dy\,dz. \end{aligned}\right\} \tag{9.84}$$

It is not hard to show that the integrals (9.84) are partial derivatives of the potential (9.83). Since the integrands in the integrals (9.83) and (9.84) are majorized by the function C/R^λ, where $\lambda = 1$ for the integral (9.83) and $\lambda = 2$ for the integrals (9.84), by Theorem 9.18 the integrals converge uniformly at every point $M(\xi, \eta, \zeta)$. Hence, by Theorem 9.17 they are continuous functions of the point $M(\xi, \eta, \zeta)$.

CHAPTER 10

FOURIER SERIES AND FOURIER INTEGRAL

It is known from linear algebra that if we choose some basis in a linear space *of finite dimension*, then any element of that space can be expanded with respect to that basis (uniquely).

Far more complicated is the question of the choice of basis and expansion with respect to it for the case of *an infinite-dimensional space.*

In the present chapter this question is studied for the case of the so-called Euclidean *infinite-dimensional* spaces and for bases of special form (the so-called *orthonormal* bases).

Particular attention is given to the study of the basis formed by the so-called *trigonometric* system in the space of all piecewise continuous functions.

A generalization of the idea of expansion of a function with respect to a basis is the expansion to be studied in the present chapter of a function into what is called *a Fourier* integral.*

Throughout this chapter the integral is understood in the Riemann sense.

10.1. ORTHONORMAL SYSTEMS AND THE GENERAL FOURIER SERIES

In the present section we shall consider an arbitrary Euclidean space *of infinite dimension***. For convenience we give the definition of a Euclidean space

Definition 1. A linear space R is said to be Euclidean *if the following two requirements are met:*

(1) *a rule is known by which any two elements f and g of R are assigned a number called* a scalar product *of those elements and designated* (f, g);

(2) *the rule satisfies the following axioms:*

1°. $(f, g) = (g, f)$ *(commutative property).*

---

* Joseph Fourier (1768-1830) is a French mathematician.

** A linear space is said to be *infinite-dimensional* if there is any preassigned number of linearly independent elements in it.

2°. $(f+g, h) = (f, h) + (g, h)$ (*distributivity*).
3°. $(\lambda f, g) = \lambda (f, g)$ *for any real number* λ.
4°. $(f, f) > 0$ *if* $f \neq 0$*, $(f, f) = 0$ *if* $f = 0$.

A classical example of an infinite-dimensional Euclidean space is *the space of all functions piecewise continuous on some closed interval* $a \leqslant x \leqslant b$.

Let us agree throughout this chapter to mean by a function $f(x)$ piecewise continuous on a closed interval $[a, b]$ a function such that is continuous everywhere on $[a, b]$, except possibly a finite number of points x_i ($i = 1, 2, \ldots, n$) at which it has a discontinuity of the first kind, and satisfies at every discontinuity point x_i the condition

$$f(x_i) = \frac{f(x_i - 0) + f(x_i + 0)}{2}. \tag{10.1}$$

Thus throughout this chapter we require that *a piecewise continuous function* $f(x)$ *should satisfy condition* (10.1) *at every discontinuity point* x_i, *i.e. should equal a half-sum of the right- and left-hand limiting values.* Note that a condition of the type (10.1) is automatically true at each continuity point of $f(x)$.

A scalar product of any two elements $f(x)$ and $g(x)$ of the space of all functions piecewise continuous on $a \leqslant x \leqslant b$ will be defined as follows:

$$(f, g) = \int_a^b f(x)\, g(x)\, dx. \tag{10.2}$$

There is no doubt that the integral (10.2) of a product of two piecewise continuous functions exists. It is easy to verify that Axioms 1° to 4° are valid for the scalar product (10.2). Axiom 1° is obvious. Axioms 2° and 3° follow from the linear properties of the integral.

We prove the validity of Axiom 4°. Since it is obvious that always $(f, f) = \int_a^b f^2(x)\, dx \geqslant 0$ it suffices to establish that from $(f, f) = \int_a^b f^2(x)\, dx = 0$ it follows that $f(x) \equiv 0$, i.e. is the zero element of the space under study. Since $f(x)$ is piecewise continuous on $[a, b]$, this interval falls into a finite number of subintervals $[x_{i-1}, x_i]$ on each of which $f(x)$ is continuous**.

---

* 0 stands for the zero element of a linear space.

** The values of $f(x)$ at the end points x_{i-1} and x_i of every interval $[x_{i-1}, x_i]$ are set equal to the limiting values $f(x_{i-1} + 0)$ and $f(x_i - 0)$ respectively.

From $\int_a^b f^2(x)\,dx = 0$ it follows that for every subinterval $[x_{i-1}, x_i]$ too

$$\int_{x_{i-1}}^{x_i} f^2(x)\,dx = 0. \tag{10.3}$$

But from (10.3) and from the continuity of $f^2(x)$ on $[x_{i-1}, x_i]$ it follows that $f(x) \equiv 0$ on $[x_{i-1}, x_i]$*.

Since the last equation relates to every subinterval and relation (10.1) is true at discontinuity points, $f(x) \equiv 0$ throughout $[a, b]$. The validity of Axiom 4° is established.

This proves that the space of all functions piecewise continuous on $[a, b]$ is a Euclidean space with scalar product (10.2).

We establish the following general property of any Euclidean space.

Theorem 10.1. In any Euclidean space, for any two elements f and g we have the inequality

$$(f, g)^2 \leqslant (f, f)\cdot(g, g) \tag{10.4}$$

which is called the Cauchy-Buniakowski inequality.

Proof. For any real number λ

$$(\lambda f - g, \lambda f - g) \geqslant 0.$$

By virtue of Axioms 1° to 4° the last inequality may be rewritten as

$$\lambda^2\cdot(f, f) - 2\lambda\,(f, g) + (g, g) \geqslant 0.$$

A necessary and sufficient condition of nonnegativeness of the last square trinomial is the nonpositiveness of its discriminant, i.e. the inequality

$$(f, g)^2 - (f, f)\cdot(g, g) \leqslant 0. \tag{10.5}$$

This immediately yields (10.4). Thus the theorem is proved.

Our next task is to introduce in the Euclidean space under study the concept of *norm* of every element.

But first we recall the definition of a normed linear space.

Definition 2. A linear space R is said to be normed *if the following two requirements are met:*

(1) *a rule is known by which each element f of R is assigned a real number called the norm of the element and designated* $\|f\|$;

(2) *the rule satisfies the following axioms:*

---

* For it has been proved in Section 10.6 of [1] that if a function is continuous, nonnegative and is not identically zero on a given closed interval, then the integral of that function over the given interval is greater than zero.

1°. $\|f\| > 0$ *if* $f \neq 0$, $\|f\| = 0$ *if* $f = 0$.
2°. $\|\lambda f\| = |\lambda|\cdot\|f\|$ *for any element* f *and any real number* λ.
3°. *For any two elements* f *and* g *we have the inequality*

$$\|f+g\| \leqslant \|f\| + \|g\| \tag{10.6}$$

called **the triangle inequality** (*or* **Minkowski inequality**).

Theorem 10.2. Any Euclidean space is a normed space if the norm of any element f *is defined in it by the equation*

$$\|f\| = \sqrt{(f, f)}. \tag{10.7}$$

Proof. It suffices to show that for the norm defined by relation (10.7) Axioms 1° to 3° of Definition 2 are true.

The validity of Axiom 1° follows immediately from Axiom 4° for a scalar product. The validity of Axiom 2° also follows almost immediately from Axioms 1° and 3° for a scalar product.

It remains to show the validity of Axiom 3°, i.e. of inequality (10.6). We shall rely on the Cauchy-Buniakowski inequality (10.4) which we rewrite as

$$|(f, g)| \leqslant \sqrt{(f, f)}\cdot\sqrt{(g, g)}.$$

Using the last inequality, Axioms 1° to 4° for a scalar product and the definition of the norm (10.7) we get

$$\|f+g\| = \sqrt{(f+g, f+g)} = \sqrt{(f, f)+2(f, g)+(g, g)} \leqslant$$
$$\leqslant \sqrt{(f, f)+2\sqrt{(f, f)}\cdot\sqrt{(g, g)}+(g, g)} =$$
$$= \sqrt{[\sqrt{(f, f)}+\sqrt{(g, g)}]^2} = \sqrt{(f, f)}+\sqrt{(g, g)} = \|f\|+\|g\|.$$

Thus the theorem is proved.

Remark. There are certainly more ways than one to introducing a scalar product (and norm) in every Euclidean space. It is sufficient for our purpose in what follows that there is at least one method of introducing a scalar product in the Euclidean space under consideration. On fixing that method, in what follows we shall always define the norm of a Euclidean space under consideration by relation (10.7). Thus, in the space of all functions piecewise continuous on $[a, b]$ (according to (10.2)) the norm is defined by the equation

$$\|f\| = \sqrt{\int_a^b f^2(x)\,dx}, \tag{10.8}$$

and the triangle inequality (10.6) has the form

$$\sqrt{\int_a^b [f(x)+g(x)]^2\,dx} \leqslant \sqrt{\int_a^b f^2(x)\,dx} + \sqrt{\int_a^b g^2(x)\,dx}. \tag{10.9}$$

We now introduce the concept of *orthogonal* elements of a given Euclidean space.

Definition 3. Two elements of a Euclidean space, f and g, are said to be orthogonal *if a scalar product (f, g) of these elements is equal to zero.*

Consider in an arbitrary infinite-dimensional Euclidean space R some sequence of elements

$$\psi_1, \psi_2, \ldots, \psi_n, \ldots . \tag{10.10}$$

Definition 4. A sequence (10.10) *is said to be* an orthonormal system *if the elements of that sequence are mutually orthogonal and have a norm equal to unity.*

A classical example of an orthonormal system in the space of all functions piecewise continuous on $-\pi \leqslant x \leqslant \pi$ is the so-called *trigonometric system*

$$\frac{1}{\sqrt{2\pi}}, \ \frac{\cos x}{\sqrt{\pi}}, \ \frac{\sin x}{\sqrt{\pi}}, \ \ldots, \ \frac{\cos nx}{\sqrt{\pi}}, \ \frac{\sin nx}{\sqrt{\pi}}, \ \ldots . \tag{10.11}$$

It is easy for the reader to verify that all the functions (10.11) are mutually orthogonal (in the sense of a scalar product (10.2) taken for $a = -\pi$, $b = \pi$) and that the norm of each of these functions (defined by equation (10.7) for $a = -\pi$, $b = \pi$) is equal to unity.

In pure and applied mathematics there occur many and various orthonormal (on corresponding sets) systems of functions.

Here are some examples of such systems.

1°. Polynomials defined by the equation

$$P_n(x) = \frac{1}{2^n \cdot n!} \frac{d^n[(x^2-1)^n]}{dx^n} \quad (n = 0, 1, 2, \ldots)$$

are generally called *Legendre polynomials.*

It is easy to see that the functions

$$\psi_n(x) = \sqrt{\frac{2n+1}{2}} \cdot P_n(x) \quad (n = 0, 1, 2, \ldots)$$

formed with the aid of these polynomials constitute an orthonormal (on $-1 \leqslant x \leqslant 1$) system of functions.

2°. Polynomials defined by the equations $T_0(x) = 1$, $T_n(x) = 2^{1-n} \cos n \times (\arccos x)$, with $n = 1, 2, \ldots$, are called *Chebyshev polynomials.* Of all n-degree polynomials with coefficients of x^n equal to unity the Chebyshev polynomial $T_n(x)$ has the smallest absolute maximum on $-1 \leqslant x \leqslant 1$. It is possible to prove that the functions

$$\psi_0(x) = \frac{1}{\sqrt{\pi} \cdot \sqrt[4]{1-x^2}}, \quad \psi_n(x) = \frac{2^{n-\frac{1}{2}} \cdot T_n(x)}{\sqrt{\pi} \cdot \sqrt[4]{1-x^2}} \quad (n = 1, 2, \ldots)$$

obtained with the aid of Chebyshev polynomials form a system orthonormal on $-1 \leqslant x \leqslant 1$.

3°. Of frequent use in probability theory is the so-called *Rademacher* system*

$\psi_n(x) = \varphi(2^n \cdot x) \quad (n = 0, 1, 2, \ldots),$

where $\varphi(t) = \mathrm{sgn}(\sin 2\pi t)$.
It can be proved that this system is orthonormal on a closed interval $0 \leqslant x \leqslant 1$.

4°. Of use in a number of investigations in function theory is the so-called *Haar** system* which is orthonormal on $0 \leqslant x \leqslant 1$. Elements of this system $\chi_{(n)}^{k}(x)$ are defined for all $n = 0, 1, \ldots$ and for all k taking on values of $1, 2, 4, \ldots, 2^n$. They are of the form

$$\chi_n^{(k)} = (x) \begin{cases} \sqrt{2^n} \text{ when } \dfrac{2k-2}{2^{n+1}} \leqslant x < \dfrac{2k-1}{2^{n+1}}, \\ -\sqrt{2^n} \text{ when } \dfrac{2k-1}{2^{n+1}} < x \leqslant \dfrac{2k}{2^{n+1}}, \\ 0 \text{ at the other points of } [0, 1]. \end{cases}$$

Every Haar function is a step of the same form as the function $\sqrt{2^n}\,\mathrm{sgn}\, x$ on a closed interval $[-2^{-(n+1)}, 2^{-(n+1)}]$. For every fixed n increasing k moves that step to the right. Everywhere outside the corresponding step every Haar function is identically zero.

Let an arbitrary orthonormal system of elements $\{\psi_k\}$ be given in an arbitrary infinite-dimensional Euclidean space R. Consider any element f of R.

Definition 5. The **Fourier series** *of an element f with respect to an orthonormal system $\{\psi_k\}$ is a series of the form*

$$\sum_{k=1}^{\infty} f_k \psi_k, \tag{10.12}$$

where f_k are constant numbers called **Fourier coefficients** *of the element f and defined by the equations*

$f_k = (f, \psi_k), \quad k = 1, 2, \ldots$

It is natural to call a finite sum

$$S_n = \sum_{k=1}^{n} f_k \psi_k \tag{10.13}$$

the nth *partial sum* of a Fourier series (10.12).

Consider along with the nth partial sum (10.13) an arbitrary linear combination of the first n elements of an orthonormal system $\{\psi_k\}$

$$\sum_{k=1}^{n} C_k \psi_k \tag{10.14}$$

with any constant numbers $C_1, C_2, \ldots, C_n$.

We shall show what distinguishes the nth partial sum of a Fourier series (10.13) from all the other sums (10.14).

* H. Rademacher (b. 1892) is a German mathematician.
** Alfred Haar (1885-1933) is a Hungarian mathematician.

Let us agree to call $\| f - g \|$ *a deviation of g from f* (in the norm of a given Euclidean space).

The following *main* theorem holds.

Theorem 10.3. Of all the sums of the form (10.14) *it is the nth partial sum* (10.13) *of the Fourier series of an element f that has the smallest deviation from f in the norm of a given Euclidean space.*

Proof. Taking into account the orthonormality of $\{\psi_k\}$ and using the axioms of a scalar product we may write

$$\Big\| \sum_{k=1}^{n} C_k\psi_k - f \Big\|^2 = \Big(\sum_{k=1}^{n} C_k\psi_k - f,\ \sum_{k=1}^{n} C_k\psi_k - f \Big) =$$

$$= \sum_{k=1}^{n} C_k^2 (\psi_k, \psi_k) - 2\sum_{k=1}^{n} C_k (f, \psi_k) + (f, f) =$$

$$= \sum_{k=1}^{n} C_k^2 - 2\sum_{k=1}^{n} C_k f_k + \| f \|^2 = \sum_{k=1}^{n} (C_k - f_k)^2 - \sum_{k=1}^{n} f_k^2 + \| f \|^2.$$

Thus

$$\Big\| \sum_{k=1}^{n} C_k\psi_k - f \Big\|^2 = \sum_{k=1}^{n} (C_k - f_k)^2 + \| f \|^2 - \sum_{k=1}^{n} f_k^2. \tag{10.15}$$

The left-hand side of (10.15) involves a squared deviation of the sum (10.14) from f (in the norm of a given Euclidean space). It follows from the form of the right-hand side of (10.15) that that squared deviation is the smallest for $C_k = f_k$ (for the first sum on the right of (10.15) vanishes and the other terms on the right of (10.15) are independent of C_k). Thus the theorem is proved.

Corollary 1. For an arbitrary element f of a given Euclidean space and any orthonormal system $\{\psi_k\}$ *under an arbitrary choice of constants* C_k *for any n*

$$\| f \|^2 - \sum_{k=1}^{n} f_k \leqslant \Big\| \sum_{k=1}^{n} C_k\psi_k - f \Big\|^2. \tag{10.16}$$

Inequality (10.16) is an immediate consequence of the identity (10.15).

Corollary 2. For an arbitrary element f of a given Euclidean space, any orthonormal system $\{\psi_k\}$ *and any n*

$$\Big\| \sum_{k=1}^{n} f_k\psi_k - f \Big\|^2 = \| f \|^2 - \sum_{k=1}^{n} f_k^2. \tag{10.17}$$

This is often called the Bessel* identity.

To prove (10.17) it suffices to set $C_k = f_k$ in (10.15).

* Friedrich Wilhelm Bessel (1784-1846) is a German astronomer and mathematician.

Theorem 10.4. For any element f of a given Euclidean space and any orthonormal system $\{\psi_k\}$ we have the following inequality:

$$\sum_{k=1}^{\infty} f_k^2 \leqslant \|f\|^2 \tag{10.18}$$

called **the Bessel inequality.**

Proof. From the nonnegativeness of the left-hand side of (10.17) it follows that for *any* n

$$\sum_{k=1}^{n} f_k^2 \leqslant \|f\|^2. \tag{10.19}$$

But this means that the series of nonnegative terms on the left of (10.18) possesses a bounded sequence of partial sums and is therefore convergent. Proceeding in inequality (10.19) to the limit as $n \to \infty$ (see Theorem 3.13 in [1]) we obtain inequality (10.18). Thus the theorem is proved.

As an illustration consider the space of all functions piecewise continuous on a closed interval $-\pi \leqslant x \leqslant \pi$ and, in that space, a Fourier series with respect to the trigonometric system (10.11) (it is customary to call that series *Fourier trigonometric series*). For any function $f(x)$ piecewise continuous on $-\pi \leqslant x \leqslant \pi$ that Fourier series has the form

$$\bar{f}_0 \frac{1}{\sqrt{2\pi}} + \sum_{k=1}^{\infty} \left(\bar{f}_k \frac{\cos kx}{\sqrt{\pi}} + \bar{\bar{f}}_k \frac{\sin kx}{\sqrt{\pi}} \right), \tag{10.20}$$

where Fourier coefficients $\bar{f}_k$ and $\bar{\bar{f}}_k$ are defined by the formulas

$$\bar{f}_0 = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x)\, dx,$$

$$\bar{f}_k = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x) \cos kx\, dx, \quad \bar{\bar{f}}_k = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x) \sin kx\, dx$$

$(k = 1, 2, \ldots)$.

The Bessel inequality true for any function $f(x)$ piecewise continuous on $-\pi \leqslant x \leqslant \pi$ has the form

$$\bar{f}_0^2 + \sum_{k=1}^{\infty} (\bar{f}_k^2 + \bar{\bar{f}}_k^2) \leqslant \int_{-\pi}^{\pi} f^2(x)\, dx. \tag{10.21}$$

In this case the deviation of $f(x)$ from $g(x)$ in the norm is equal to the so-called *root-mean-square* (or *standard*) *deviation*

$$\|f-g\| = \sqrt{\int_{-\pi}^{\pi} [f(x)-g(x)]^2\,dx}. \tag{10.22}$$

In the theory of Fourier trigonometric series, however, a somewhat different notation for both the Fourier series (10.20) and the Bessel inequality (10.21) is used. It is the Fourier trigonometric series (10.20) that is usually written as

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \tag{10.20'}$$

where

$$\left.\begin{aligned} a_0 &= \frac{2\bar{f}_0}{\sqrt{2\pi}} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx, \\ a_k &= \frac{\bar{f}_k}{\sqrt{\pi}} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx, \\ b_k &= \frac{\bar{\bar{f}}_k}{\sqrt{\pi}} = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx \\ (k &= 1, 2, \ldots). \end{aligned}\right\} \tag{10.23}$$

With this notation the Bessel inequality (10.21) becomes

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \leqslant \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)\,dx. \tag{10.21'}$$

Remark. It follows from the Bessel inequality (10.21') that for any function $f(x)$ piecewise continuous on $-\pi \leqslant x \leqslant \pi$ the quantities a_k and b_k (called *Fourier trigonometric coefficients* of $f(x)$) tend to zero as $k \to \infty$ (by the necessary condition for convergence of the series on the left of (10.21')).

10.2. CLOSED AND COMPLETE ORTHONORMAL SYSTEMS

As in the preceding section we shall consider an arbitrary orthonormal system $\{\psi_k\}$ in any infinite-dimensional Euclidean space R.

Definition 1. An orthonormal system $\{\psi_k\}$ *is said to be* closed *if for any element* f *of a given Euclidean space* R *and for any positive number* ε *there is a linear combination (10.14) of a finite number of elements of* $\{\psi_k\}$ *such that its deviation from* f *(in the norm of* R*) is less than* ε.

In other words, a system $\{\psi_k\}$ is said to be closed if any element f of a given Euclidean space R can be approximated in the norm of that space to an unlimited precision by linear combinations of a finite number of elements of $\{\psi_k\}$.

Remark 1. We drop the question of whether there are closed orthonormal systems in any Euclidean space. Note that in Chapter 11 we study an important subclass of Euclidean spaces, the so-called *Hilbert* spaces, and establish that there are closed orthonormal systems in every such space.

Theorem 10.5. If an orthonormal system $\{\psi_k\}$ is closed, then for any element f of a given Euclidean space the Bessel inequality (10.18) *goes over into an exact equation*

$$\sum_{k=1}^{\infty} f_k^2 = \|f\|^2 \tag{10.24}$$

called **Parseval's* formula.**

Proof. Fix an arbitrary element f of the Euclidean space under consideration and an arbitrary positive number ε. Since the system $\{\psi_k\}$ is closed, there is n and numbers $C_1, C_2, \ldots, C_n$ such that the square of the norm on the right of (10.16) is less than ε. By virtue of (10.16) this means that for an arbitrary $\varepsilon > 0$ there is n for which

$$\|f\|^2 - \sum_{k=1}^{n} f_k^2 < \varepsilon. \tag{10.25}$$

For integers greater than that n inequality (10.25) is clearly true, for the sum on the left of (10.25) may only increase as n increases.

So we have proved that for an arbitrary $\varepsilon > 0$ there is n beginning with which inequality (10.25) is true.

In conjunction with inequality (10.19) this means that the series $\sum_{k=1}^{\infty} f_k^2$ converges to the sum $\|f\|^2$. Thus the theorem is proved.

Theorem 10.6. If an orthonormal system $\{\psi_k\}$ is closed, then whatever the element f the Fourier series of that element converges to it in the norm of the space under consideration, i.e.

$$\lim_{n\to\infty} \left\| \sum_{k=1}^{n} f_k\psi_k - f \right\| = 0. \tag{10.26}$$

Proof. The statement of this theorem follows immediately from equation (10.17) and from the preceding theorem.

Remark 2. In the space of all functions piecewise continuous on a closed interval $-\pi \leqslant x \leqslant \pi$ convergence in the norm (10.26)

* M. Parseval (d. 1836) is a French mathematician.

turns into convergence *in the mean* on that interval (see Section 1.2.3). Thus, if the closure of the trigonometric system (10.11) is proved, Theorem 10.6 will state that for any function $f(x)$ piecewise continuous on $-\pi \leqslant x \leqslant \pi$ the Fourier trigonometric series of that function converges to it on $-\pi \leqslant x \leqslant \pi$ in the mean.

Definition 2. An orthonormal system $\{\psi_k\}$ *is said to be* complete *if besides the zero element there is no other element f of a given Euclidean space that would be orthogonal to all the elements* ψ_k *of* $\{\psi_k\}$.

In other words, a system $\{\psi_k\}$ is said to be complete if any element f orthogonal to all the elements ψ_k of $\{\psi_k\}$ is a zero element.

Theorem 10.7. Any closed orthonormal system $\{\psi_k\}$ *is complete.*

Proof. Let a system $\{\psi_k\}$ be closed and let f be any element of a given Euclidean space orthogonal to all the elements ψ_k of $\{\psi_k\}$. Then all Fourier coefficients f_k of f with respect to $\{\psi_k\}$ are zero and therefore, by Parseval's formula (10.24), $\| f \| = 0$ too. The last equation (by Axiom 1° for the norm) implies that $f = 0$. Thus the theorem is proved.

Remark 3. We have proved that in an arbitrary Euclidean space the closure of an orthonormal system implies its completeness. In Chapter 11 an example is given showing that in an arbitrary Euclidean space the completeness of an orthonormal system does not in general imply its closure. It is also proved there that for a very important class of Euclidean spaces, the so-called Hilbert spaces, the completeness of an orthonormal system is equivalent to its closure.

Theorem 10.8. For any complete (and clearly for any closed) orthonormal system $\{\psi_k\}$ *two distinct elements f and g of a given Euclidean space cannot have the same Fourier series.*

Proof. If all Fourier coefficients of f and g coincided, then all Fourier coefficients of the difference $f - g$ would be zero, i.e. the difference $f - g$ would be orthogonal to all the elements ψ_k of a complete system $\{\psi_k\}$. But this would mean that $f - g$ is a zero element, i.e. that the elements f and g coincide. Thus the theorem is proved.

This completes our discussion of the general Fourier series with respect to an arbitrary orthonormal system in any Euclidean space.

Out next goal is to make a detailed study of the Fourier series with respect to the trigonometric system (10.11).

10.3. THE COMPLETENESS OF THE TRIGONOMETRIC SYSTEM. COROLLARIES

10.3.1. Uniform approximation to a continuous function by trigonometric polynomials. We shall establish here the closure (and therefore completeness) of the trigonometric system (10.11) in the space of all functions piecewise continuous on a closed interval $-\pi \leqslant x \leqslant \pi$. But before proving the closure of the trigonometric system

we shall establish an important theorem on uniform approximation to a continuous function by the so-called trigonometric polynomials.

A trigonometric polynomial is an arbitrary linear combination of any finite number of elements of the trigonometric system (10.11), i.e. an expression of the form

$$T(x)=\overline{C}_0+\sum_{k=1}^{n}(\overline{C}_k\cos kx+\overline{\overline{C}}_k\sin kx),$$

where n is any integer and $\overline{C}_k$ and $\overline{\overline{C}}_k$ $(k = 1, 2, \ldots, n)$ are arbitrary constant reals.

Note two quite elementary *statements*:

1°. *If $P(x)$ is any algebraic polynomial of arbitrary degree n, then* $P(\cos x)$ *and* $P(\sin x)$ *are trigonometric polynomials.*

2°. *If $T(x)$ is a trigonometric polynomial, then either of the expressions $T(x)\cdot\sin x$ and $T(x)\cdot\sin^2 x$ is a trigonometric polynomial.*

Both statements follow from the fact that the product of two (and therefore of any finite number of) trigonometric functions* of x reduces to a linear combination of a finite number of trigonometric functions of independent variables of the type kx (see this for yourself).

In the theory of Fourier trigonometric series an important role is played by the concept of *periodic* function.

A function $f(x)$ is said to be a periodic *function with period T if: (1) $f(x)$ is defined for all real x; (2) for any real x*

$$f(x+T)=f(x).$$

This equation is usually called *periodicity condition*. Consideration of periodic functions results from the study of various oscillatory processes.

Note that all the elements of the trigonometric system (10.11) are periodic functions with period 2π.

The following *main* theorem is true.

Theorem 10.9 (Weierstrass). If a function $f(x)$ is continuous on a closed interval $[-\pi, \pi]$ and satisfies the condition $f(-\pi) = f(\pi)$, then it can be uniformly approximated on $[-\pi, \pi]$ by trigonometric polynomials, i.e. for this function $f(x)$ and for any positive number ε there is a trigonometric polynomial $T(x)$ such that for all x in $[-\pi, \pi]$ at once

$$|f(x)-T(x)|<\varepsilon. \tag{10.27}$$

Proof. For convenience we carry out the proof in two steps.

1°. First suppose additionally that $f(x)$ is *even*, i.e. satisfies the condition $f(-x) = f(x)$ for any x in $[-\pi, \pi]$.

* In this case we mean cosine and sine by trigonometric functions.

By the theorem on the continuity of the composite function $y = f(x)$, where $x = \arccos t$ (see Section 4.7 of [1]), the function $F(t) = f(\arccos t)$ is a continuous function of the independent variable t on a closed interval $-1 \leqslant t \leqslant 1$. Therefore by the Weierstrass theorem for algebraic polynomials (see Theorem 1.18) given any $\varepsilon > 0$ we can find an algebraic polynomial $P(t)$ such that $|f(\arccos t) - P(t)| < \varepsilon$ at once for all t in $-1 \leqslant t \leqslant 1$.

On letting $t = \cos x$ we get

$$|f(x) - P(\cos x)| < \varepsilon \tag{10.28}$$

at once for all x in $0 \leqslant x \leqslant \pi$.

Since both functions $f(x)$ and $P(\cos x)$ are *even*, inequality (10.28) is true for all x in $-\pi \leqslant x \leqslant 0$, too. Thus (10.28) is true for all x in $-\pi \leqslant x \leqslant \pi$ and since (by Statement 1° above) $P(\cos x)$ is a trigonometric polynomial, this proves the theorem for an even function $f(x)$.

Now note that a function $f(x)$ satisfying the hypotheses of the theorem being proved can be, periodically with period 2π, extended to the whole infinite straight line $-\infty < x < \infty$, so that the extended function will be continuous at each point x of the line. If $f(x)$ is extended in this way, then since $P(\cos x)$ is also a periodic function of period 2π, we find that *for an* even *function $f(x)$ inequality* (10.28) *is true everywhere on an infinite straight line $-\infty < x < \infty$.*

2°. Now let $f(x)$ be an entirely arbitrary function satisfying the hypotheses of the theorem. We extend it periodically with period 2π to the whole infinite straight line and form with the aid of it the following two *even* functions:

$$f_1(x) = \frac{f(x) + f(-x)}{2} \tag{10.29}$$

$$f_2(x) = \frac{f(x) - f(-x)}{2} \sin x. \tag{10.30}$$

By what was proved in 1°, for any $\varepsilon > 0$ there are trigonometric polynomials $T_1(x)$ and $T_2(x)$ such that everywhere on the infinite straight line

$$|f_1(x) - T_1(x)| < \varepsilon/4, \quad |f_2(x) - T_2(x)| < \varepsilon/4,$$

and therefore

$$|f_1(x)\sin^2 x - T_1(x)\sin^2 x| < \varepsilon/4,$$
$$|f_2(x)\sin x - T_2(x)\sin x| < \varepsilon/4.$$

Adding together the last two inequalities, considering that the modulus or absolute value of the sum of two quantities is not greater than the sum of their moduli and taking into consideration equations (10.29) and (10.30) we see that everywhere on the infinite straight

line

$$|f(x)\sin^2 x - T_3(x)| < \varepsilon/2, \tag{10.31}$$

where $T_3(x)$ is a trigonometric polynomial equal to $T_3(x) = T_1(x) \times \sin^2 x + T_2(x) \sin x$.

Instead of $f(x)$ in the above discussion we can take a function $f(x + \pi/2)$*. In full analogy with (10.31) we shall have that for $f(x + \pi/2)$ there is a trigonometric polynomial $T_4(x)$ such that everywhere on an infinite straight line

$$|f(x+\pi/2)\sin^2 x - T_4(x)| < \varepsilon/2. \tag{10.32}$$

Replacing in (10.32) x by $x - \pi/2$ and denoting by $T_5(x)$ a trigonometric polynomial of the form $T_5(x) = T_4(x - \pi/2)$ we see that everywhere on an infinite straight line

$$|f(x)\cos^2 x - T_5(x)| < \varepsilon/2. \tag{10.33}$$

Finally, additing together inequalities (10.31) and (10.33) and denoting by $T(x)$ a trigonometric polynomial of the form $T(x) = T_4(x) + T_5(x)$ we see that everywhere on the infinite straight line inequality (10.27) is true. Thus the theorem is proved.

Remark. Each of the conditions, (1) the continuity of $f(x)$ on $[-\pi, \pi]$, (2) the equality of $f(-\pi)$ and $f(\pi)$ in value is a *necessary* condition for $f(x)$ to be uniformly approximated on $-\pi \leqslant x \leqslant \pi$ by trigonometric polynomials.

In other words, the Weirstrass theorem can be restated as follows: *for a function $f(x)$ to be approximated uniformly on a closed interval $[-\pi, \pi]$ by trigonometric polynomials, it is necessary and sufficient that $f(x)$ should be continuous on $[-\pi, \pi]$ and satisfy the condition $f(-\pi) = f(\pi)$.*

Sufficiency makes the content of Theorem 10.9.

We shall prove the *necessity*. Let there be a sequence of trigonometric polynomials $\{T_n(x)\}$ converging uniformly on $[-\pi, \pi]$ to $f(x)$. Since every function $T_n(x)$ is continuous on $[-\pi, \pi]$, so is $f(x)$, by Theorem 1.8. Given any $\varepsilon > 0$ we can find a polynomial $T_n(x)$ such that $|f(x) - T_n(x)| < \varepsilon/2$ for every x in $[-\pi, \pi]$. Therefore

$$|f(-\pi) - T_n(-\pi)| < \varepsilon/2, \quad |f(\pi) - T_n(\pi)| < \varepsilon/2.$$

From the last two inequalities and from the equation $T_n(-\pi) = T_n(\pi)$ following from the periodicity (with period 2π) condition we conclude that $|f(-\pi) - f(\pi)| < \varepsilon$ whence $f(-\pi) = f(\pi)$ (by the arbitrariness of $\varepsilon > 0$).

10.3.2. The proof of the closure of the trigonometric system. Relying on the Weierstrass theorem we prove the following *main* theorem.

* For this function satisfies the same conditions as the extended function $f(x)$ does.

Theorem 10.10. The trigonometric system (10.11) *is closed**, i.e. *for any function $f(x)$ piecewise continuous on a closed interval $[-\pi, \pi]$ and any positive number ε there is a trigonometric polynomial $T(x)$ such that*

$$\|f(x)-T(x)\|=\sqrt{\int_{-\pi}^{\pi}[f(x)-T(x)]^2\,dx}<\varepsilon. \tag{10.34}$$

Proof. First of all note that for any function $f(x)$ piecewise continuous on $[-\pi, \pi]$ and for any $\varepsilon > 0$ there is a function $F(x)$ continuous on $[-\pi, \pi]$, satisfying the condition $F(-\pi) = F(\pi)$ and such that

$$\|f(x)-F(x)\|=\sqrt{\int_{-\pi}^{\pi}[f(x)-F(x)]^2\,dx}<\varepsilon/2. \tag{10.35}$$

Indeed, it suffices to take the function $F(x)$ to coincide with $f(x)$ everywhere except sufficiently small neighbourhoods of the discontinuity points of $f(x)$ and of the point $x = \pi$ and to take $F(x)$ to be a linear function in those neighbourhoods so that $F(x)$ is continuous throughout $[-\pi, \pi]$ and satisfies $F(-\pi) = F(\pi)$.

Since a piecewise continuous function and the linear function completing it are bounded, choosing those neighbourhoods of the discontinuity points of $f(x)$ and of the point $x = \pi$ to be sufficiently small ensures that inequality (10.35) holds.

By the Weierstrass theorem 10.9, given a function $F(x)$ we can find a trigonometric polynomial $T(x)$ such that for every x in $[-\pi, \pi]$

$$|F(x) - T(x)| \leqslant \varepsilon/2\sqrt{2\pi}. \tag{10.36}$$

From (10.36) we conclude that

$$\|F(x)-T(x)\| = \sqrt{\int_{-\pi}^{\pi}[F(x)-T(x)]^2\,dx}\leqslant\varepsilon/2. \tag{10.37}$$

From (10.35) and (10.37) and from the triangle inequality for the norms

$$\|f(x) - T(x)\| \leqslant \|f(x) - F(x)\| + \|F(x) - T(x)\|$$

we obtain inequality (10.34). Thus the theorem is proved.

Remark 1. From Theorems 10.10 and 10.7 it follows immediately that *the trigonometric system* (10.11) *is complete*. From this in turn it follows that *the system* $\left\{\sqrt{\frac{2}{\pi}}\sin nx\right\}$

* And therefore (by Theorem 10.7) *complete*.

($n=1, 2, \ldots$) *is complete on the set of all functions piecewise continuous on a closed interval* $[0, \pi]$ (*or respectively on* $[-\pi, 0]$). Indeed, any function $f(x)$ piecewise continuous on $[0, \pi]$ and orthogonal on $[0, \pi]$ to all the elements of the system $\left\{\sqrt{\frac{2}{\pi}}\sin nx\right\}$ turns out to be orthogonal on $[-\pi, \pi]$ to *all* the elements of the trigonometric system (10.11) after an odd extension to $[-\pi, 0]$. By the completeness of the system (10.11) that function is zero on $[-\pi, \pi]$ and therefore on $[0, \pi]$. Quite similarly it can be proved that *the system* $\frac{1}{\sqrt{\pi}}$, $\sqrt{\frac{2}{\pi}}\cos nx$ ($n=1, 2, \ldots$) *is complete on the set of all functions piecewise continuous on* $[0, \pi]$ (or respectively on $[-\pi, 0]$).

Remark 2. It can be shown that of the orthonormal systems indicated in Section 10.1 those formed using Legendre polynomials, Chebyshev polynomials and Haare functions are closed systems and the Rademacher system is not.

10.3.3. Corollaries.

Corollary 1. For any function $f(x)$ piecewise continuous on a closed interval $[-\pi, \pi]$ **Parseval's formula**

$$\frac{a_0^2}{2}+\sum_{k=1}^{\infty}(a_k^2+b_k^2)=\frac{1}{\pi}\int_{-\pi}^{\pi}f^2(x)\,dx \tag{10.38}$$

holds (this follows from Theorem 10.5).

Corollary 2. The Fourier trigonometric series of any function $f(x)$ piecewise continuous on $[-\pi, \pi]$ converges to that function on $[-\pi, \pi]$ **in the mean** (this follows from Theorem 10.6 and Remark 2 to it).

Corollary 3. The Fourier trigonometric series of any function $f(x)$ piecewise continuous on $[-\pi, \pi]$ can be integrated term by term on $[-\pi, \pi]$ (this follows from Corollary 2 above and from Theorem 1.11).

Corollary 4. If two functions $f(x)$ and $g(x)$ piecewise continuous on $[-\pi, \pi]$ have the same Fourier trigonometric series, then they coincide everywhere on $[-\pi, \pi]$ (this follows from Theorem 10.8).

Corollary 5. If the Fourier trigonometric series of a function $f(x)$ piecewise continuous on $[-\pi, \pi]$ converges uniformly on some closed interval $[a, b]$ contained in $[-\pi, \pi]$ then it is to the function $f(x)$ that it converges on $[a, b]$.

Proof. Let $F(x)$ be the function to which the Fourier trigonometric series of $f(x)$ uniformly converges on $[a, b]$. We prove that $F(x)\equiv f(x)$ throughout $[a, b]$. Since uniform convergence on $[a, b]$ implies convergence in the mean on it (see Section 1.2.3), the Fourier trigonometric series of $f(x)$ converges to $F(x)$ on $[a, b]$ in the mean. This means that for an arbitrary $\varepsilon > 0$ there is n_1 beginning wi[illegible] which the nth partial sum of the Fourie[illegible]onometric series S

satisfies the inequality

$$\| F(x)-S_n(x) \| = \sqrt{\int_a^b [F(x)-S_n(x)]^2\,dx} < \varepsilon/2. \qquad (10.39)$$

On the other hand, by Corollary 2 the sequence $S_n(x)$ converges to $f(x)$ in the mean throughout $[-\pi, \pi]$ and therefore on $[a, b]$ too, i.e. for the fixed arbitrary $\varepsilon > 0$ there is n_ε beginning with which

$$\| S_n(x)-f(x) \| = \sqrt{\int_a^b [S_n(x)-f(x)]^2\,dx} < \varepsilon/2. \qquad (10.40)$$

From (10.39) and (10.40) and from the triangle inequality

$$\| F(x) - f(x) \| \leqslant \| F(x) - S_n(x) \| + \| S_n(x) - f(x) \|$$

it follows that $\| F(x) - f(x) \| < \varepsilon$. From the last inequality and from the arbitrariness of $\varepsilon > 0$ it follows that $\| F(x) - f(x) \| = 0$ and from this on the basis of Axiom 1° for the norm we conclude that $F(x) - f(x)$ is *the zero element* of a space of functions piecewise continuous on $[a, b]$, i.e. a function identically zero on $[a, b]$. Thus Corollary 5 is proved.

Remark 1. Of course, in Corollary 5 the closed interval $[a, b]$ may coincide with the entire interval $[-\pi, \pi]$, i.e. *from the uniform convergence of the Fourier series of $f(x)$ throughout $[-\pi, \pi]$ it follows that it is to $f(x)$ that that series converges on that interval.*

Remark 2. Quite similar corollaries will be true for a Fourier series with respect to any other closed orthonormal system in the space of functions piecewise continuous on an arbitrary interval $[a, b]$ with scalar product (10.2) and norm (10.8). The orthonormal systems associated with Legendre and Chebyshev polynomials and the Haare system (see Section 10.1) may serve as examples of such systems.

10.4. THE SIMPLEST CONDITIONS FOR THE UNIFORM CONVERGENCE AND TERM-BY-TERM DIFFERENTIATION OF A FOURIER TRIGONOMETRIC SERIES

10.4.1. Introductory remarks. Of great importance in mathematical physics and in a number of other branches of mathematics is the question of under what conditions the Fourier trigonometric series of a function $f(x)$ converges (to that function) at a given point x of a closed interval $[-\pi, \pi]$.

It was known as far back as the late XIX century that there are functions, continuous on $[-\pi, \pi]$ and satisfying the condition $f(-\pi) = f(\pi)$, whose Fourier trigonometric series diverge at a prea-

signed point of $[-\pi, \pi]$ (or diverge even on an infinite set of points of $[-\pi, \pi]$ everywhere dense on that interval)*.

Thus the continuity alone of $f(x)$ on $[-\pi, \pi]$ without any additional conditions fails to ensure not only the uniform convergence of the Fourier trigonometric series of that function but even the convergence of the series at a preassigned point of $[-\pi, \pi]$.

In this and the next section we shall explore what requirements should be added to the continuity of $f(x)$ (or introduced instead of the continuity of $f(x)$) to ensure the convergence of the Fourier trigonometric series of $f(x)$ at a given point as well as its uniform convergence throughout $[-\pi, \pi]$ or some part of it.

Yet another question arises in the study of the convergence of a Fourier trigonometric series: must the Fourier trigonometric series of any function $f(x)$ piecewise continuous (or even strictly continuous) on $[-\pi, \pi]$ converge *at least at one point of the interval*?

A positive answer to that question was obtained only in 1966.

It is a consequence of the fundamental theorem, proved in 1966 by L. Carleson**, that has solved N. N. Luzin's*** famous problem formulated as early as 1914: *The Fourier trigonometric series of any function $f(x)$ for which there is an integral $\int_{-\pi}^{\pi} f^2(x)\,dx$ understood in the Lebesgue sense converges to that function almost everywhere on $[-\pi, \pi]$*****.

From Carleson's theorem it follows that the Fourier series of any function $f(x)$ integrable on $[-\pi, \pi]$ in the proper Riemann sense, as well as that of any piecewise continuous function $f(x)$, converges almost everywhere on $[-\pi, \pi]$ to $f(x)$ (since there is an integral $\int_{-\pi}^{\pi} f^2(x)\,dx$ in the Riemann sense and therefore in the Lebesgue sense for such a function).

Note that if a function $f(x)$ is integrable on a closed interval $[-\pi, \pi]$ not in the Riemann sense but only in the Lebesgue sense, then the Fourier trigonometric series of that function may fail to

---

* The earliest example of such a function was constructed by the French mathematician P. du Bois-Reymond in 1876.

** Lennart Carleson is a modern Swedish mathematician. A full proof of Carleson's theorem can be found in *Selected problems on exceptional sets* by Lennart Carleson (Russian translation in *Matematika*, Vol. II, No. 4, 1967, pp. 113-132).

*** Nikolay Nikolayevich Luzin (1883-1950) is a Soviet mathematician, the founder of the modern Moscow school of function theory. The statement of Luzin's problem solved by Carleson and of his other problems can be found in N.N. Luzin. *The Integral and Trigonometric Series*, Moscow, Leningrad, Gostekhizdat (1951).

**** For the definition of the integral in the Lebesgue sense and of the convergence almost everywhere on a given closed interval see Chapter 8.

converge at any point of $[-\pi, \pi]$. The Soviet mathematician A. N. Kolmogorov* was the first to construct in 1923 an example of a function $f(x)$ integrable on $[-\pi, \pi]$ in the Lebesgue sense, with an everywhere divergent Fourier trigonometric series.

10.4.2. The simplest conditions for absolute and uniform convergence of a Fourier trigonometric series. Let us agree to use the following terminology:

Definition 1. We shall say that a function $f(x)$ has a piecewise continuous derivative *on a closed interval $[a, b]$ if the derivative $f'(x)$ exists and is continuous everywhere on $[a, b]$ except possibly for a finite number of points at each of which the function $f'(x)$ has finite right- and left-hand limiting values**.*

Definition 2. We shall say that a function $f(x)$ has a piecewise continuous derivative of order $n \geqslant 1$ on a closed interval $[a, b]$ if the function $f^{(n-1)}(x)$ has on that interval a piecewise continuous derivative in the sense of Definition 1.

The following *main* theorem is true.

Theorem 10.11. If a function $f(x)$ is continuous on a closed interval $[-\pi, \pi]$, has on it a piecewise continuous derivative and satisfies the condition $f(-\pi) = f(\pi)$, then the Fourier trigonometric series of $f(x)$ converges to that function uniformly on $[-\pi, \pi]$. Moreover, the series made up of the absolute values of the terms of the Fourier trigonometric series of $f(x)$ converges uniformly on $[-\pi, \pi]$.

Proof. It suffices to prove that the series made up of the absolute values of the terms of the Fourier trigonometric series of $f(x)$

$$\frac{|a_0|}{2} + \sum_{k=1}^{\infty} \{|a_k \cos kx| + |b_k \sin kx|\} \tag{10.41}$$

converges uniformly on $[-\pi, \pi]$, for this will imply both the uniform convergence on $[-\pi, \pi]$ of the Fourier trigonometric series itself of $f(x)$ and that it is to the function $f(x)$ that that series converges (by Corollary 5 of Section 10.3.3).

By virtue of the Weierstrass test (see Theorem 1.4), to prove uniform convergence on $[-\pi, \pi]$ of the series (10.41) it suffices to establish the convergence of the number series

$$\sum_{k=1}^{\infty} \{|a_k| + |b_k|\} \tag{10.42}$$

---

* A construction of A.N. Kolmogorov's example can be found on pages 412 to 421 of the book: N.N. Bari, *Trigonometric Series*, Moscow, Fizmatgiz (1961).

** The function $f'(x)$ may turn out to be undefined at a finite number of points of $[a, b]$. We specify it arbitrarily at those points (for instance, we set it equal to the half-sum of the right- and left-hand limiting values).

majorizing it. Denote by α_k and β_k the Fourier trigonometric coefficients of the function $f'(x)$ specifying the function arbitrarily in a finite number of points at which there is not derivative of $f(x)$*.

Integrating by parts and considering that $f(x)$ is continuous throughout $[-\pi, \pi]$ and satisfies the relation $f(-\pi) = f(\pi)$ we obtain the following relations connecting the Fourier trigonometric coefficients of the function $f'(x)$ and those of the function $f(x)$ itself:**

$$\alpha_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\cos kx\,dx = k\cdot\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx = k\cdot b_k,$$

$$\beta_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\sin kx\,dx = -k\cdot\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx = -k\cdot a_k.$$

Thus

$$|a_k| + |b_k| = \frac{|\alpha_k|}{k} + \frac{|\beta_k|}{k},$$

and, to prove the convergence of the series (10.42), it suffices to establish the convergence of the series

$$\sum_{k=1}^{\infty}\left\{\frac{|\alpha_k|}{k} + \frac{|\beta_k|}{k}\right\}. \tag{10.43}$$

The convergence of the series (10.43) follows from the elementary inequalities***

$$\frac{|\alpha_k|}{k} \leqslant \frac{1}{2}\left(\alpha_k^2 + \frac{1}{k^2}\right),$$

$$\frac{|\beta_k|}{k} \leqslant \frac{1}{2}\left(\beta_k^2 + \frac{1}{k^2}\right) \tag{10.44}$$

and from the convergence of the series

$$\sum_{k=1}^{\infty}(\alpha_k^2 + \beta_k^2), \quad \sum_{k=1}^{\infty}\frac{1}{k^2}, \tag{10.45}$$

---

* For instance, we may set the function $f'(x)$ at those points equal to a half-sum of the right- and left-hand limiting values.

** When integrating by parts one should divide $[-\pi, \pi]$ into a finite number of subintervals having no interior points in common on each of which the derivative $f'(x)$ is continuous and, taking an integration by parts formula for each of the subintervals, consider that when summing the integrals over all subintervals all the substitutions vanish (because of the continuity of $f(x)$ throughout $[-\pi, \pi]$ and the conditions $f(-\pi) = f(\pi)$).

*** We proceed from the elementary inequality $|a|\cdot|b| \leqslant \frac{1}{2}(a^2+b^2)$ following from the nonnegativity of $(|a|-|b|)^2$.

the first of them converging by virtue of Parseval's formula for a piecewise continuous function $f'(x)$ and the second by virtue of the Cauchy-Maclaurin integral test (see Section 13.2 of [1]). Thus the theorem is proved.

Remark. If a function $f(x)$ satisfying the hypotheses of Theorem 10.11 is periodically (with period 2π) extended to the whole infinite straight line, then Theorem 10.11 will state convergence of a Fourier trigonometric series to the function thus extended that is *uniform throughout the infinite straight line.*

10.4.3. The simplest conditions for term-by-term differentiation of a Fourier trigonometric series. We shall first prove the following lemma on the order of Fourier trigonometric coefficients.

Lemma 1. Let a function $f(x)$ and all of its derivatives up to some order m (m being a nonnegative integer) *be continuous on a closed interval $[-\pi, \pi]$ and satisfy the conditions*

$$\left.\begin{aligned} &f(-\pi) = f(\pi),\\ &f'(-\pi) = f'(\pi),\\ &\dots\dots\dots\dots\\ &f^{(m)}(-\pi) = f^{(m)}(\pi). \end{aligned}\right\} \tag{10.46}$$

Also let $f(x)$ have on $[-\pi, \pi]$ a piecewise continuous derivative of order $(m+1)$. Then the following series converges:

$$\sum_{k=1}^{\infty} k^m \{|a_k| + |b_k|\}, \tag{10.47}$$

where a_k and b_k are the Fourier trigonometric coefficients of $f(x)$.

Proof. Denote by α_k and β_k the Fourier trigonometric coefficients of the function $f^{(m+1)}(x)$, specifying the function arbitrarily in a finite number of points at which the function $f(x)$ has no derivative of order $(m+1)$. Integrating the expressions for α_k and β_k $(m+1)$ times by parts and considering the continuity throughout $[-\pi, \pi]$ of the function $f(x)$ itself and of all its derivatives up to order m and taking into account relations (10.46) we establish the following relation of the Fourier trigonometric coefficients of the function $f^{(m+1)}(x)$ to the function $f(x)$ itself*:

$$|\alpha_k| + |\beta_k| = k^{m+1}\{|a_k| + |b_k|\}.$$

Thus

$$k^m\{|a_k| + |b_k|\} = \frac{|\alpha_k|}{k} + \frac{|\beta_k|}{k},$$

* When integrating by parts one should divide the closed interval $[-\pi, \pi]$ into a finite number of subintervals having no interior points in common on each of which $f^{(m+1)}(x)$ is continuous and consider that summing the integrals over all the subintervals makes all substitutions vanish.

and the convergence of the series (10.47) follows from the elementary inequalities (10.44) and from the convergence of the series (10.45), the first of them converging by virtue of Parseval's formula for a piecewise continuous function $f^{(m+1)}(x)$ and the second by virtue of the Cauchy-Maclaurin test. Thus the lemma is proved.

An immediate consequence of Lemma 1 is the following theorem.

Theorem 10.12. Let a function $f(x)$ satisfy the same conditions as in Lemma 1, with $m \geqslant 1$. Then the Fourier trigonometric series of $f(x)$ can be differentiated term by term m times on $[-\pi, \pi]$.

Proof. Let s be any of the numbers $1, 2, \ldots, m$. As a result of an s-fold term-by-term differentiation of the Fourier trigonometric series of $f(x)$ we obtain the series

$$\sum_{k=1}^{\infty} k^s \left\{a_k \cos\left(kx - \frac{\pi s}{2}\right) + b_k \sin\left(kx - \frac{\pi s}{2}\right)\right\}. \tag{10.48}$$

Note that for every x in $[-\pi, \pi]$ both the original Fourier trigonometric series and the series (10.48) (with any $s = 1, 2, \ldots, m$) can be majorized by the convergent number series (10.47). By the Weierstrass test (see Theorem 1.4) both the original Fourier trigonometric series and each of the series (10.48) (with $s = 1, 2, \ldots, m$) converge uniformly on $[-\pi, \pi]$ and this (by virtue of Theorem 1.9) ensures the possibility of an m-fold term-by-term differentiation of the original Fourier series. Thus the theorem is proved.

10.5. MORE PRECISE CONDITIONS FOR UNIFORM CONVERGENCE AND CONDITIONS FOR CONVERGENCE AT A GIVEN POINT

10.5.1. The modulus of continuity of a function. Hölder classes. We shall begin with explaining the concepts characterizing the smoothness of the functions under study and with defining the classes of functions in terms of which we shall formulate the conditions for convergence of a Fourier trigonometric series.

Let a function $f(x)$ be defined and continuous on a closed interval $[a, b]$.

Definition 1. For every $\delta > 0$ the modulus of continuity *of $f(x)$ on $[a, b]$ is the supremum of the absolute value of the difference $|f(x') - f(x'')|$ on the set of all x' and x'' of $[a, b]$ satisfying the condition $|x' - x''| < \delta$.*

We shall denote the modulus of continuity of $f(x)$ on $[a, b]$ by $\omega(\delta, f)$. So by definition*

$$\omega(\delta, f) = \sup_{\substack{|x'-x''|<\delta \\ x', x'' \in [a, b]}} |f(x') - f(x'')|.$$

* Recall that the symbol $\in$ stands for "belongs to, is in", so that the notation $x', x'' \in [a, b]$ designates that the points x' and x'' are in a closed interval $[a, b]$.

It is immediate from the Cantor theorem (see Theorem 10.2 in [1]) that *the modulus of continuity* $\omega(\delta, f)$ *of any function* $f(x)$ *continuous on* $[a, b]$ *tends to zero as* $\delta \to 0$*.

However for an arbitrary function $f(x)$ only continuous on $[a, b]$ nothing in general can be said *about the order* of its modulus of continuity $\omega(\delta, f)$ with respect to a small δ.

We now show that *if a function* $f(x)$ *is differentiable on* $[a, b]$ *and if its derivative* $f'(x)$ *is bounded on* $[a, b]$, *then the modulus of continuity of* $f(x)$ *on that interval,* $\omega(\delta, f)$, *has an order* $\omega(\delta, f) = O(\delta)$**.

Indeed, it follows from the Lagrange*** theorem that for any points x' and x'' in $[a, b]$ there is a point ξ contained between x' and x'' such that

$$|f(x') - f(x'')| = |f'(\xi)|\cdot|x' - x''|. \tag{10.49}$$

Since the derivative $f'(x)$ is bounded on $[a, b]$ there is a constant M such that for every x in the interval $|f'(x)| \leqslant M$ and therefore $|f'(\xi)| \leqslant M$. From the last inequality and from (10.49) we conclude that $|f(x') - f(x'')| \leqslant M\delta$ for every x' and x'' in $[a, b]$ satisfying $|x' - x''| < \delta$. But this means that $\omega(\delta, f) \leqslant M\delta$, i.e. $\omega(\delta, f) = O(\delta)$.

Let α be any real number in the half-open interval $0 < \alpha \leqslant 1$.

Definition 2. We shall say that a function $f(x)$ *belongs on a closed interval* $[a, b]$ *to a Hölder class* C^α *with index* α $(0 < \alpha \leqslant 1)$ *if the modulus of continuity of* $f(x)$ *on* $[a, b]$ *has an order* $\omega(\delta, f) = O(\delta^\alpha)$.

To designate that $f(x)$ belongs to a Hölder class C^α on $[a, b]$ one usually uses the notation: $f(x) \in C^\alpha[a, b]$.

Note at once that if a function $f(x)$ is differentiable on $[a, b]$ and its derivative is bounded on that interval, then $f(x)$ a fortiori belongs on $[a, b]$ to the Hölder class C^1**** (this statement follows immediately from the relation $\omega(\delta, f) = O(\delta)$ proved above).

Remark. Let $f(x) \in C^\alpha[a, b]$. The supremum of the fraction $\dfrac{|f(x')-f(x'')|}{|x'-x''|^\alpha}$ on the set of all x' and x'' of $[a, b]$ not equal to each other is called *the Hölder constant* (or *the Hölder coefficient*) of $f(x)$ (on $[a, b]$). The sum of the Hölder constant of $f(x)$ on $[a, b]$ and the supremum $|f(x)|$ on that interval is called *the Hölder norm* of $|f(x)|$ on $[a, b]$ and designated $\|f\|_{C^\alpha[a, b]}$.

---

* For (by the Cantor theorem) given any $\varepsilon > 0$ we can find $\delta > 0$ such that $|f(x') - f(x'')| < \varepsilon$ for every x' and x'' in $[a, b]$ satisfying the condition $|x' - x''| < \delta$.

** Recall that the symbol $\alpha = O(\delta)$ was introduced in Chapters 3 and 4 of [1] and denotes the existence of a constant M such that $|\alpha| \leqslant M\delta$.

*** See Theorem 8.12 in [1].

**** The Hölder class C^1 corresponding to the value $\alpha = 1$ is often called *the Lipschitz class*.

Example. The function $f(x)=\sqrt{x}$ belongs on a closed interval $[0, 1]$ to the class $C^{1/2}$, for $|f(x')-f(x'')|=\sqrt{x'-x''}\times\times\frac{\sqrt{x'-x''}}{\sqrt{x'}+\sqrt{x''}}\leqslant\sqrt{x'-x''}$ for any x' and x'' in $[0, 1]$ related by the condition $x'>x''$ (*the Hölder constant* that is the supremum on $[0, 1]$ of the fraction $\frac{\sqrt{x'-x''}}{\sqrt{x'}+\sqrt{x''}}$, being equal to unity and *the Hölder norm* to two).

10.5.2. An expression for the partial sum of a Fourier trigonometric series. Let $f(x)$ be an arbitrary function piecewise continuous on a closed interval $[-\pi, \pi]$. We shall extend that function periodically (with period 2π) to the whole infinite straight line*. Denote by $S_n(x, f)$ a partial sum of the Fourier trigonometric series of $f(x)$ at a point x equal to

$$S_n(x, f)=\frac{a_0}{2}+\sum_{k=1}^{n}(a_k\cos kx+b_k\sin kx). \tag{10.50}$$

Inserting into the right-hand side of (10.50) the values of the Fourier coefficients**

$$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(y)\,dy,$$

$$a_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(y)\cos ky\,dy,\quad b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(y)\sin ky\,dy\quad (k=1, 2, \ldots)$$

and taking into account the linear properties of the integral we get

$$S_n(x, f)=\frac{1}{\pi}\int_{-\pi}^{\pi}f(y)\left[\frac{1}{2}+\sum_{k=1}^{n}(\cos ky\cos kx+\sin ky\sin kx)\right]dy=$$

$$=\frac{1}{\pi}\int_{-\pi}^{\pi}f(y)\left[\frac{1}{2}+\sum_{k=1}^{n}\cos k(y-x)\right]dy$$

* By the convention adopted as far back as in Section 10.1 a piecewise continuous function $f(x)$ must have at every point x a value equal to a half-sum of the right- and left-hand limiting values. For this property to take place and for $f(x)$ to be periodically (with period 2π) extended to the whole infinite straight line we must require that the relation $f(\pi)=f(-\pi)=(1/2)[f(-\pi+0)+f(\pi-0)]$ should hold for the extended function. In other words, we shall say that a function $f(x)$ defined on an infinite straight line is *a periodic extension* of a function $f(x)$ piecewise continuous on a closed interval $[-\pi, \pi]$ if both functions coincide on the interval $-\pi<x<\pi$ and if the function $f(x)$ defined on the infinite straight line satisfies the periodicity condition $f(x+2\pi)=f(x)$ and the condition $f(\pi)=f(-\pi)=(1/2)[f(-\pi+0)+f(\pi-0)]$.

** See formulas (10.23).

for any point x of an infinite straight line. On making in the last integral a change of variable $y = t + x$ we arrive at the following expression

$$S_n(x, f) = \frac{1}{\pi} \int_{-\pi-x}^{\pi-x} f(x+t) \left[\frac{1}{2} + \sum_{k=1}^{n} \cos kt\right] dt. \tag{10.51}$$

Note now that since either of the functions $f(x+t)$ and $\left[\frac{1}{2} + \sum_{k=1}^{n} \cos kt\right]$ is a periodic function of the variable t with period 2π, the entire integrand in (10.51) (denote it briefly by $F(t)$) is a periodic function of t with period 2π. Also note that integration in (10.51) is over $[-\pi-x, \pi-x]$ of length equal to 2π, i.e. equal to the period of the integrand. We use the following elementary *statement: if $F(t)$ is a periodic function of period 2π integrable over any finite closed interval, then all integrals of that function over any of the closed intervals of length equal to the period 2π are equal to one another, i.e. for any x*

$$\int_{-\pi-x}^{\pi-x} F(t)\,dt = \int_{-\pi}^{\pi} F(t)\,dt\,^{*}. \tag{10.52}$$

Equation (10.52) allows us to rewrite formula (10.51) as follows:

$$S_n(x, f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+t) \left[\frac{1}{2} + \sum_{k=1}^{n} \cos kt\right] dt. \tag{10.53}$$

Compute the sum in the square brackets in (10.53). To do this note that for any k and any value of t

$$2 \sin \frac{t}{2} \cos kt = \sin\left(k + \frac{1}{2}\right) t - \sin\left(k - \frac{1}{2}\right) t.$$

---

* To prove this statement it suffices, using the additive property, to represent $\int_{-\pi-x}^{\pi-x} F(t)\,dt$ as the sum of three integrals

$$\int_{-\pi-x}^{-\pi} F(t)\,dt + \int_{-\pi}^{\pi} F(t)\,dt + \int_{\pi}^{\pi-x} F(t)\,dt$$

and to note that with the aid of the periodicity condition $F(t) = F(t + 2\pi)$ and the change of variable $t = y - 2\pi$ the first of the three integrals is reduced to the third taken with the minus sign. Indeed,

$$\int_{-\pi-x}^{-\pi} F(t)\,dt = \int_{-\pi-x}^{-\pi} F(t+2\pi)\,dt = \int_{\pi-x}^{\pi} F(y)\,dt = -\int_{\pi}^{\pi-x} F(y)\,dy.$$

Summing this equation over all k equal to 1, 2, . . ., n we get

$$2\sin\frac{t}{2}\cdot\sum_{k=1}^{n}\cdot\cos kt=\sin\left(n+\frac{1}{2}\right)t-\sin\frac{t}{2}.$$

Hence

$$2\sin\frac{t}{2}\left[\frac{1}{2}+\sum_{k=1}^{n}\cos kt\right]=\sin\left(n+\frac{1}{2}\right)t$$

and therefore

$$\left[\frac{1}{2}+\sum_{k=1}^{n}\cos kt\right]=\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}. \tag{10.54}$$

Substituting (10.54) in (10.53) we finally obtain the following expression for a partial sum of a Fourier trigonometric series:

$$S_n(x,\ f)=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x+t)\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt \tag{10.55}$$

true at any point x of an infinite straight line.

Remark. From formula (10.55) and from the fact that all partial sums $S_n(x,\ 1)$ of the function $f(x)\equiv 1$ are equal to unity* we get

$$1=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt. \tag{10.56}$$

10.5.3. The integral modulus of continuity of a function. Let a function $f(x)$ be integrable (in the sense of an improper Riemann integral) on a closed interval $[-\pi,\ \pi]$. We extend that function periodically (with period 2π) to the whole infinite straight line.

Definition. *For any δ of the half-open interval $0<\delta\leqslant 2\pi$ the* **integral modulus of continuity** *of a function $f(x)$ on a closed interval $[-\pi,\ \pi]$ is the supremum of the integral*

$$\int_{-\pi}^{\pi}|f(t+u)-f(t)|\,dt$$

on the set of all numbers u satisfying the condition $|u|\leqslant\delta$.

We shall denote the integral modulus of continuity of a function $f(x)$ on a closed interval $[-\pi,\ \pi]$ by $I(\delta,\ f)$.

* For the quantity (10.55) for the funct[illegible]) $\equiv 1$ is equal [illegible] (10.50), where $a_0=2$, $a_k=b_k=0$, $k=1$[illegible]

So by definition

$$I(\delta, f) = \sup_{|u| \leqslant \delta} \int_{-\pi}^{\pi} |f(t+u) - f(t)|\, dt.$$

The following statement is true.

Lemma 2. If a function $f(x)$ is piecewise continuous on a closed interval $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line, then the integral modulus of continuity of the function on that interval, $I(\delta, f)$, tends to zero as $\delta \to 0$.

Proof. Fix an arbitrary $\varepsilon > 0$. According to Theorem 10.10 (on the closure of the trigonometric system) for a function $f(x)$ there is a trigonometric polynomial $T(x)$ such that

$$\|f - T\| = \sqrt{\int_{-\pi}^{\pi} [f(t) - T(t)]^2\, dt} < \varepsilon/3\sqrt{2\pi},$$

and therefore on the basis of the Cauchy-Buniakowski inequality*

$$\int_{-\pi}^{\pi} |f(t) - T(t)|\, dt \leqslant \sqrt{\int_{-\pi}^{\pi} [f(t) - T(t)]^2\, dt \int_{-\pi}^{\pi} dt} < \varepsilon/3. \qquad (10.57)$$

From inequality (10.57) and from the fact that $f(t)$ and $T(t)$ are periodic functions of period 2π we conclude that for any number u

$$\int_{-\pi}^{\pi} |f(t+u) - T(t+u)|\, dt < \varepsilon/3. \qquad (10.58)$$

Since the absolute value of the sum of three quantities is at most the sum of the absolute values of those quantities, for any number u

$$\int_{-\pi}^{\pi} |f(t+u) - f(t)|\, dt \leqslant \int_{-\pi}^{\pi} |f(t+u) - T(t+u)|\, dt +$$

$$+ \int_{-\pi}^{\pi} |T(t+u) - T(t)|\, dt + \int_{-\pi}^{\pi} |T(t) - f(t)|\, dt. \qquad (10.59)$$

Now it remains to note that by virtue of the continuity of a trigonometric polynomial and the Cantor theorem (see Theorem 10.2 in [1]) for the fixed $\varepsilon > 0$ there is $\delta > 0$ such that for $|u| \leqslant \delta$ and every t in $[-\pi, \pi]$

$$|T(t+u) - T(t)| < \varepsilon/6\pi,$$

* See inequality (1.33).

and therefore

$$\int_{-\pi}^{\pi} |T(t+u)-T(t)|\,dt \leqslant \varepsilon/3. \tag{10.60}$$

Comparing (10.59) with inequalities (10.57), (10.58), and (10.60) we get

$$\int_{-\pi}^{\pi} |f(t+u)-f(t)|\,dt < \varepsilon \tag{10.61}$$

for every u for which $|u| \leqslant \delta$. Thus the lemma is proved.

Remark to Lemma 2. It is easy to see that *the integral modulus of continuity $I(\delta, f)$ tends to zero as $\delta \to 0$ not only for any piecewise continuous function $f(x)$ but also for any $f(x)$ integrable (in the proper Riemann sense) on a closed interval $[-\pi, \pi]$*. To prove this fix an arbitrary $\varepsilon > 0$ and note that by virtue of the integrability of $f(t)$ on $-\pi \leqslant t \leqslant \pi$ there is $\delta_0 > 0$ such that for *any* subdivision of $[-\pi, \pi]$ into subintervals of length smaller than δ_0 the difference between the upper and lower sums of $f(t)$ is less than $\varepsilon/4$. Fix some subdivision T of $[-\pi, \pi]$ into subintervals of *equal* length $\delta < \delta_0$. From the fact that $f(t)$ is a periodic function it follows that for any $|u| \leqslant \delta$ and for the fixed subdivision T of $-\pi \leqslant t \leqslant \pi$ the difference between the upper and lower sums of the function $f(t+u)$ (with δ sufficiently small) is at least less than $\varepsilon/2$. But from this it follows that for the fixed subdivision T the difference between the upper and lower sums of the function $[f(t+u)-f(t)]$, with any $|u| \leqslant \delta$, is less than $\varepsilon/4 + \varepsilon/2 = {}^3/_4\,\varepsilon$. Denote for the fixed subdivision T the upper and lower sums of $[f(t+u)-f(t)]$ respectively by S and s and the upper and lower sums of the function $|f(t+u)-f(t)|$ respectively by $\overline{S}$ and $\overline{s}$. It has been established in Section 10.5 of [1] that for any subdivision the upper and lower sums S and s of the function itself and the upper and lower sums $\overline{S}$ and $\overline{s}$ of the absolute value of that function are connected by the relation $\overline{S} - \overline{s} \leqslant S - s$. Thus, for the fixed subdivision T, $\overline{S} - \overline{s} < 3\varepsilon/4$. But this means that for the fixed subdivision T the difference between *any* integral sum of the function $|f(t+u)-f(t)|$ and the integral $\int_{-\pi}^{\pi} |f(t+u)-f(t)|\,dt$ is less than the number $3\varepsilon/4$. If we choose in that integral sum all the intermediate points ξ_k in the centre of the corresponding subintervals of length δ and require that the number u should satisfy the inequality $|u| < \delta/2$, then both points ξ_k and $\xi_k + u$ will be in the kth subinterval and therefore the difference $|f(\xi_k+u)-f(\xi_k)|$ will not exceed the oscillation $M_k - m_k$ of $f(t)$ on the kth subinterval*. But then the entire integral sum is not greater than the sum $\sum (M_k - m_k)\,\Delta t_k$ equal to the difference between the upper and lower sums of $f(t)$ of a subdivision T, i.e. is not greater than the number $\varepsilon/4$. It follows

* By M_k and m_k we denote the supremum and infimum of $f(t)$ on the kth subinterval.

that for $|u| < \delta/2$ the integral $\int_{-\pi}^{\pi} |f(t-u) - f(t)|\,dt$ is not greater than the number ε, which proves that $I(\delta, f)$ tends to zero as $\delta \to 0$.

We now derive from Lemma 2 a number of corollaries important for what follows.

Corollary 1. If a function $f(t)$ is piecewise continuous on a closed interval $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line and x is any fixed point of $[-\pi, \pi]$, then for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$\int_{-\pi}^{\pi} |f(x+t+u) - f(x+t)|\,dt < \varepsilon, \tag{10.62}$$

with $|u| < \delta$.

Proof. On making in the integral on the left of (10.62) a change of variable $\tau = x + t$

$$\int_{-\pi}^{\pi} |f(x+t+u) - f(x+t)|\,dt = \int_{-\pi+x}^{\pi+x} |f(\tau+u) - f(\tau)|\,d\tau$$

and observing that (by virtue of equation (10.52))

$$\int_{-\pi+x}^{\pi+x} |f(\tau+u) - f(\tau)|\,d\tau = \int_{-\pi}^{\pi} |f(\tau+u) - f(\tau)|\,d\tau,$$

we see that inequality (10.62) is a consequence of (10.61).

Corollary 2. If each of the functions $f(t)$ and $g(t)$ is piecewise continuous on $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line, then the function

$$I(x) = \int_{-\pi}^{\pi} f(x+t)\,g(t)\,dt$$

is a continuous function of x on a closed interval $-\pi \leqslant x \leqslant \pi$.

Proof. Let x be any point of $[-\pi, \pi]$. Then

$$I(x+u) - I(x) = \int_{-\pi}^{\pi} [f(x+t+u) - f(x+t)]\,g(t)\,dt.$$

and since the function $g(t)$ piecewise continuous on $[-\pi, \pi]$ satisfies on that interval the boundedness condition $|g(t)| \leqslant M$, we have

$$|I(x+u) - I(x)| \leqslant M \int_{-\pi}^{\pi} |f(x+t+u) - f(x+t)|\,dt$$

and therefore in view of (10.62) for any $\varepsilon > 0$

$$|I(x+u) - I(x)| < \varepsilon, \quad \text{if } |u| < \delta(\varepsilon).$$

Thus the continuity of $I(x)$ at a point x is proved.

Corollary 3. If each of the functions $f(t)$ and $g(t)$ is piecewise continuous on $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line, then the Fourier trigonometric coefficients of the function $F(x, t) = f(x+t)\,g(t)$, when this function is expanded in terms of t,

$$a_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\,g(t)\cos nt\,dt, \tag{10.63}$$

$$b_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\,g(t)\sin nt\,dt \tag{10.64}$$

converge to zero (as $n \to \infty$) uniformly with respect to x on $[-\pi, \pi]$ (and therefore on the whole infinite straight line).

Proof. For any fixed point x of $[-\pi, \pi]$ the function $F(x, t) = f(x+t)\,g(t)$ is a piecewise continuous function of the independent variable t on $[-\pi, \pi]$ and therefore Parseval's formula

$$\frac{a_0^2(x)}{2} + \sum_{k=1}^{\infty}[a_k^2(x) + b_k^2(x)] = \frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x+t)\,g^2(t)\,dt \tag{10.65}$$

holds for this function*. Equation (10.65) implies the convergence of the series on its left at each fixed point x of $[-\pi, \pi]$. Since that series consists of *nonnegative* terms, to prove its uniform convergence on $[-\pi, \pi]$ it suffices by the Dini theorem** to establish that both each of the functions $a_n(x)$ and $b_n(x)$ and the sum of the series (10.65) $\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x+t)\,g^2(t)\,dt$ are continuous functions of x on $[-\pi, \pi]$, but this follows at once from the preceding corollary (it suffices to consider that the square of a piecewise continuous function is a piecewise continuous function and that $\cos nt$ and $\sin nt$, with every n fixed, are continuous functions).

Corollary 4. If each of the functions $f(t)$ and $g(t)$ is piecewise continuous on $[-\pi, \pi]$ and periodically (with period 2π) extended to the

* See Corollary 1 in Section 10.3.3.

** See Theorem 1.5 (statement in terms of series).

whole infinite straight line, then the sequence

$$c_n(x)=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\,g(t)\sin\left(n+\frac{1}{2}\right)t\,dt \qquad (10.66)$$

converges to zero uniformly with respect to x on $[-\pi, \pi]$ (and therefore on the whole infinite straight line).

Proof. It suffices to consider that

$$\sin\left(n+\frac{1}{2}\right)t=\cos nt\cdot\sin\frac{t}{2}+\sin nt\cdot\cos\frac{t}{2}$$

and apply Corollary 3, taking the function $g(t)\cdot\sin\frac{t}{2}$ instead of $g(t)$ in (10.63) and the function $g(t)\cdot\cos\frac{t}{2}$ instead of $g(t)$ in (10.64).

10.5.4. The principle of localization. We shall prove here that the question of whether the Fourier trigonometric series of a function $f(x)$ piecewise continuous on $[-\pi, \pi]$ and periodical (with period 2π) converges or diverges at a given point x_0 can be solved *only on the basis of the behaviour of the function $f(x)$ in arbitrarily small neighbourhood of x_0*. This remarkable property of a Fourier trigonometric series is generally called *the principle of localization.*

We begin by proving an important lemma.

Lemma 3 (Riemann). If a function $f(x)$ is piecewise continuous on $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line and if it vanishes on some interval $[a, b]$, then for any positive number δ smaller than $\frac{b-a}{2}$ the Fourier trigonometric series of $f(x)$ converges to zero uniformly on the interval $[a+\delta, b-\delta]$.*

Proof. Let δ be an arbitrary positive number smaller than $\frac{b-a}{2}$. The partial sum of the Fourier trigonometric series of $f(x)$ at an arbitrary point x of an infinite straight line is defined by equation (10.55). Setting

$$g(t)=\begin{cases}\dfrac{1}{2\sin\frac{t}{2}} & \text{when } \delta\leqslant|t|\leqslant\pi,\\ 0 & \text{when } |t|<\delta\end{cases} \qquad (10.67)$$

and considering that $f(x+t)$ is zero provided x is in $[a+\delta, b-\delta]$ and t is in the interval $|t|\leqslant\delta$** we may rewrite (10.55) for each

* The closed interval $[a, b]$ is quite arbitrary. In particular, it may be contained entirely in $[-\pi, \pi]$.

** By virtue of the fact that $f(x)$ is zero throughout $[a, b]$.

x of $[a+\delta,\ b-\delta]$ as

$$S_n(x, f)=\frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\, g(t) \sin\left(n+\frac{1}{2}\right) t\, dt.$$

It remains to take into consideration that by Corollary 4 of Section 10.3.3 the sequence on the right of the last equation converges to zero uniformly with respect to x on the whole infinite straight line. Thus the lemma is proved.

Immediate consequences of Lemma 3 are the following two theorems.

Theorem 10.13. *Let a function $f(x)$ be piecewise continuous on a closed interval $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line and let $[a, b]$ be some closed interval. For the Fourier trigonometric series of $f(x)$ to converge, with any positive δ smaller than $\frac{b-a}{2}$, (to $f(x)$) uniformly on the interval $[a+\delta, b-\delta]$ it is sufficient that there is a periodic (with period 2π) function $g(x)$ piecewise continuous on $[-\pi, \pi]$ that has a Fourier trigonometric series uniformly converging on $[a, b]$ and coincides on $[a, b]$ with the function $f(x)$.*

Proof. Applying Lemma 3 to the difference $[f(x)-g(x)]$ we see that the Fourier trigonometric series of the difference $[f(x)-g(x)]$ converges to zero uniformly on $[a+\delta, b-\delta]$ for any δ in the interval $0<\delta<\frac{b-a}{2}$, and from this and from the uniform convergence on $[a, b]$ of the Fourier trigonometric series of $g(x)$ we obtain the uniform convergence on $[a+\delta, b-\delta]$ of the Fourier trigonometric series of $f(x)$. The fact that *it is to* $f(x)$ that the last series converges on $[a+\delta, b-\delta]$ follows immediately from Corollary 5 of Section 10.3.3. Thus the theorem is proved.

Theorem 10.14. *Let a function $f(x)$ be piecewise continuous on a closed interval $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line and let x_0 be some point of the line. For the Fourier trigonometric series of $f(x)$ to converge at x_0 it is sufficient that there is a periodic (with period 2π) function $g(x)$ piecewise continuous on $[-\pi, \pi]$ that has a Fourier trigonometric series converging at the point x_0 and coincides with $f(x)$ in an arbitrarily small δ-neighbourhood of x_0.*

Proof. It suffices to apply Lemma 3 to the difference $[f(x)-g(x)]$ on the interval $\left[x_0-\frac{\delta}{2}, x_0+\frac{\delta}{2}\right]$ and consider that the convergence at x_0 of the trigonometric series of the functions $[f(x)-g(x)]$ and $g(x)$ implies the convergence at that point of the Fourier trigonometric series of $f(x)$ as well. Thus the theorem is proved.

Theorem 10.14 does not establish any particular sort of conditions that ensure the convergence of the Fourier trigonometric series of $f(x)$ at a point x_0. It only proves that those conditions are determined solely by the behaviour of $f(x)$ in an arbitrarily small neighbourhood of x_0 (i.e. have a *local* character).

10.5.5. Uniform convergence of a Fourier trigonometric series for functions of a Hölder class. In this and the next subsection we shall discuss more precise conditions ensuring the uniform convergence and convergence of a Fourier trigonometric series at a given point.

We prove the following *main* theorem.

Theorem 10.15. If a function $f(x)$ belongs on $[-\pi, \pi]$ to a Hölder class C^α with an arbitrary positive index α $(0 < \alpha \leqslant 1)$ and if in addition $f(-\pi) = f(\pi)$, then the Fourier trigonometric series of $f(x)$ converges (to that function) uniformly on $[-\pi, \pi]$.

Proof. We shall assume as usual that $f(x)$ is extended periodically (with period 2π) to the whole infinite straight line. The condition $f(-\pi) = f(\pi)$ ensures that the function thus extended belongs to a Hölder class C^α on the whole infinite straight line.

Let x be any point of $[-\pi, \pi]$. Multiplying both sides of (10.56) by $f(x)$ and subtracting the resulting equation from (10.55) we get

$$S_n(x, f) - f(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} [f(x+t) - f(x)] \frac{\sin\left(n - \frac{1}{2}\right)t}{2\sin\frac{t}{2}}\, dt. \qquad (10.68)$$

The condition that $f(x)$ belongs to a Hölder class C^α implies that there is M such that

$$|f(x+t) - f(x)| \leqslant M \cdot t^\alpha \qquad (10.69)$$

in any case for every x and every t in $[-\pi, \pi]$.

Fix an arbitrary $\varepsilon > 0$ and choose for it $\delta > 0$ satisfying

$$\frac{M}{\alpha} \cdot \delta^\alpha < \frac{\varepsilon}{3}. \qquad (10.70)$$

Dividing $[-\pi, \pi]$ into a sum of the closed interval $|t| \leqslant \delta$ and the set $\delta \leqslant |t| \leqslant \pi$, we make equation (10.68) take the form

$$S_n(x, f) - f(x) = \frac{1}{\pi}\int_{|t| \leqslant \delta} [f(x+t) - f(x)] \frac{\sin\left(n + \frac{1}{2}\right)t}{2\sin\frac{t}{2}}\, dt +$$

$$+ \frac{1}{\pi}\int_{\delta \leqslant |t| \leqslant \pi} f(x+t) \frac{\sin\left(n - \frac{1}{2}\right)t}{2\sin\frac{t}{2}}\, dt -$$

$$- \frac{f(x)}{\pi}\int_{\delta \leqslant |t| \leqslant \pi} \frac{\sin\left(n - \frac{1}{2}\right)t}{2\sin\frac{t}{2}}\, dt. \qquad (10.71)$$

To evaluate the first of the integrals on the right of (10.71) use inequality (10.69) and consider that $\frac{1}{2\left|\sin\frac{t}{2}\right|}\leqslant\frac{\pi}{2|t|}$ for every t in $[-\pi, \pi]$*.

We see that for any n and any x in $[-\pi, \pi]$

$$\left|\int\limits_{|t|\leqslant\delta}[f(x+t)-f(x)]\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt\right|\leqslant$$

$$\leqslant\int\limits_{|t|\leqslant\delta}|f(x+t)-f(x)|\frac{\left|\sin\left(n+\frac{1}{2}\right)t\right|}{2\left|\sin\frac{t}{2}\right|}\cdot dt\leqslant$$

$$\leqslant\frac{M\pi}{2}\int\limits_{|t|\leqslant\delta}|t|^{\alpha-1}\,dt=M\pi\int\limits_0^\delta t^{\alpha-1}\,dt=\frac{M\pi}{\alpha}\cdot\delta^\alpha.$$

From this in view of (10.70), for any n and any x in $[-\pi, \pi]$

$$\left|\frac{1}{\pi}\int\limits_{|t|\leqslant\delta}[f(x+t)-f(x)]\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt\right|<\frac{\varepsilon}{3}. \tag{10.72}$$

Using the function (10.67) piecewise continuous on the second of the integrals on the right of (10.71) can be written as

$$\frac{1}{\pi}\int\limits_{\delta\leqslant|t|\leqslant\pi}f(x+t)\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt=$$

$$=\frac{1}{\pi}\int\limits_{-\pi}^{\pi}f(x+t)\,g(t)\sin\left(n+\frac{1}{2}\right)t.$$

By Corollary 4 of Section 10.5.3 the right-hand side of the last equation converges to zero (as $n\to\infty$) uniformly with respect to x on

---

* This inequality follows immediately from the fact that the function $\frac{\sin x}{x}$ *decreases* from 1 to $2/\pi$ as x changes from 0 to $\pi/2$. The fact that $\frac{\sin x}{x}$ decreases follows in turn from the fact that $\left(\frac{\sin x}{x}\right)'=\frac{\cos x}{x^2}\times$ $\times(x-\tan x)<0$ everywhere for $0<x<\pi/2$, since $x<\tan x$ for $0<x<\pi/2$ (see Section 4.5.6 of [1]).

$[-\pi, \pi]$. For the fixed $\varepsilon > 0$ therefore there is N_1 such that

$$\left| \frac{1}{\pi} \int\limits_{\delta \leqslant |t| \leqslant \pi} f(x+t) \frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}} dt \right| < \frac{\varepsilon}{3} \tag{10.73}$$

for every $n \geqslant N_1$ and every x in $[-\pi, \pi]$.

To evaluate the last integral on the right of (10.71) notice that using the piecewise continuous function (10.67) this integral can be written as

$$\frac{f(x)}{\pi} \int\limits_{\delta \leqslant |t| \leqslant \pi} \frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}} dt = \frac{f(x)}{\pi} \int\limits_{-\pi}^{\pi} g(t) \sin\left(n+\frac{1}{2}\right) t\, dt.$$

The integral on the right of the last equation converges to zero (as $n \to \infty$) also by Corollary 4 of Section 10.5.3 (it suffices to apply that corollary to the function $f(x) \equiv 1$). Considering also that $f(x)$ is in any case bounded on $[-\pi, \pi]$ we see that for the fixed arbitrary $\varepsilon > 0$ there is N_2 such that

$$\left| \frac{f(x)}{\pi} \int\limits_{\delta \leqslant |t| \leqslant \pi} \frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}} dt \right| < \frac{\varepsilon}{3} \tag{10.74}$$

for every $n \geqslant N_2$ and every x in $[-\pi, \pi]$.

On denoting by N the largest of the two integers N_1 and N_2 we find by virtue of (10.71) to (10.74) that for the fixed arbitrary $\varepsilon > 0$ there is N such that

$$| S_n(x, f) - f(x) | < \varepsilon$$

for every $n \geqslant N$ and for every x in $[-\pi, \pi]$. Thus the theorem is proved.

Remark 1. It is obvious that under the hypotheses of Theorem 10.15 the Fourier trigonometric series converges uniformly not only on $[-\pi, \pi]$, but also *uniformly on the whole infinite straight line* (to the function which is a periodic (with period 2π) extension of $f(x)$ to the whole infinite straight line).

Remark 2. Note that in evaluating the integrals (10.73) and (10.74) we used only the piecewise continuity (and the resulting boundedness) of the function $f(x)$ on $[-\pi, \pi]$ (no use was made of the fact that $f(x)$ belongs to a Hölder class).

Remark 3. The question naturally arises as to whether it is possible to slacken in Theorem 10.15 the smoothness requirement on $f(x)$ while preserving the statement about the uniform convergence on $[-\pi, \pi]$ of the Fourier trigonometric series of $f(x)$.

Recall that the fact that $f(x)$ belongs on $[-\pi, \pi]$ to a Hölder class C^{α} means by definition that the modulus of continuity of $f(x)$ on $[-\pi, \pi]$ has an order

$$\omega(\delta, f) = O(\delta^{\alpha}).$$

We note without proof the so-called *Dini-Lipschitz theorem* which states that *for the Fourier trigonometric series of $f(x)$ to converge uniformly on $[-\pi, \pi]$ it is sufficient that $f(x)$ satisfies $f(-\pi) = f(\pi)$ and that its modulus of continuity on $[-\pi, \pi]$ has an order*

$$\omega(\delta, f) = o\left(\frac{1}{\ln 1/\delta}\right),$$

i.e. is an infinitesimal, as $\delta \to 0$, of an order higher than $\frac{1}{\ln 1/\delta}$.

The Dini-Lipschitz theorem contains a *final* condition (in terms of the modulus of continuity of a function) for uniform convergence of the Fourier trigonometric series of the function, for it is possible to construct a function $f(x)$ satisfying $f(-\pi) = f(\pi)$ with a modulus of continuity of order $O\left(\frac{1}{\ln 1/\delta}\right)$ on $[-\pi, \pi]$ and with a Fourier trigonometric series diverging on a set of points everywhere dense on $[-\pi, \pi]$*.

Under the hypotheses of Theorem 10.15, after a periodic (with period 2π) extension the function $f(x)$ was found to belong to a Hölder class C^{α} *on the whole infinite straight line.* The question naturally arises as to the behaviour of the Fourier trigonometric series of a function $f(x)$ belonging to a Hölder class C^{α} *only on some closed interval* $[a, b]$ and satisfying but the ordinary piecewise-continuity requirement everywhere outside it.

The answer to this question is given by the following theorem.

Theorem 10.16. *Let a function $f(x)$ be piecewise continuous on $[-\pi, \pi]$ and periodically (with period 2π) extended to the whole infinite straight line. Suppose further that on some closed interval $[a, b]$ of length smaller than 2π it belongs to a Hölder class C^{α} with an arbitrary positive index α $(0 < \alpha \leqslant 1)$. Then for any δ in the interval $0 < \delta < \frac{b-a}{2}$ the Fourier trigonometric series of $f(x)$ converges (to that function) uniformly on the closed interval $[a + \delta, b - \delta]$.*

Proof. Construct a function $g(x)$ that coincides with $f(x)$ on $[a, b]$, is a linear function of the form $Ax + B$ on $[b, a + 2\pi]$ turning into $f(b)$ when $x = b$ and into $f(a)$ when $x = a + 2\pi$** and that is periodically (with period 2π) extended from a closed interval $[a, a + 2\pi]$ to the whole infinite straight line (in Fig. 10.1 the heavy line ...

* A proof of the Dini-Lipschitz theorem and a construction of the ex... just mentioned can be found in the book: A. Zygmund, *Trigonometric S...*, vol. 1, Cambridge, 1959.

** The condition that the function $Ax + B$ turns into $f(b)$ when ... and into $f(a)$ when $x = a + 2\pi$ uniquely defines the constants ...

$$A = \frac{f(a) - f(b)}{a + 2\pi - b}, \quad B = \frac{(a + 2\pi) f(b) - bf(a)}{a + 2\pi - b}.$$

presents the graph of $f(x)$ and the broken line is the graph of a function $g(x)$ constructed from $f(x)$).

It is obvious that the constructed function $g(x)$ satisfies the condition $g(-\pi) = g(\pi)$ and belongs to a Hölder class C^{α} (having the same positive index α as $f(x)$) on the whole infinite straight line*. By Theorem 10.15 and Remark 1 the Fourier trigonometric series of $g(x)$ converges uniformly on the whole infinite straight line and therefore by Theorem 10.13 the Fourier trigonometric series

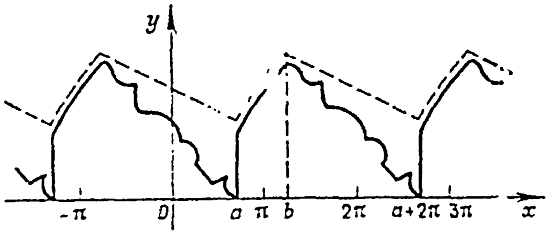


Fig. 10.1

of $f(x)$ converges (to that function) uniformly on $[a+\delta, b-\delta]$ for any δ in the interval $0 < \delta < \frac{b-a}{2}$. Thus the theorem is proved.

Remark 4. The statement of Theorem 10.16 remains true for a closed interval $[a, b]$ of length *equal to* 2π (i.e. for the case $b = a + 2\pi$), but then in proving the theorem, on fixing an arbitrary δ in the interval $0 < \delta < \pi$, we should take the function $g(x)$ to coincide with $f(x)$ on $\left[a+\frac{\delta}{2}, a+2\pi-\frac{\delta}{2}\right]$, to be linear on $\left[a+2\pi-\frac{\delta}{2}, a+2\pi+\frac{\delta}{2}\right]$ and periodically (with period 2π) extended from $\left[a+\frac{\delta}{2}, a+2\pi+\frac{\delta}{2}\right]$ to the whole infinite straight line. But if $[a, b]$ is of length *greater than* 2π, then from $f(x)$ belonging to a Hölder class C^{α} on such an interval and from the periodicity condition of $f(x)$ (with period 2π) it follows that $f(x)$ belongs to a class C^{α} on the whole infinite straight line, i.e. in this case we arrive at Theorem 10.15.

10.5.6. On the convergence of the Fourier trigonometric series of a piecewise Hölder function.

Definition 1. We shall say that a function $f(x)$ *is piecewise Hölder on a closed interval* $[a, b]$ *if it is piecewise continuous on* $[a, b]$ *and if* $[a, b]$ *can be divided by means of a finite number of points* $a = x_0 <$

---

* It suffices to consider that $g(x)$ is everywhere continuous and that a linear function has a bounded derivative and belongs therefore to a Hölder class C^{α} for any $\alpha \leqslant 1$.

$< x_1 < x_2 < \ldots < x_n = b$ *into subintervals* $[x_{k-1}, x_k]$ $(k = 1, 2, \ldots, n)$ *on each of which that function belongs to a Hölder class* C^{α_k} *with some positive index* α_k $(0 < \alpha_k \leqslant 1)$; *when defining the Hölder class on a subinterval* $[x_{k-1}, x_k]$ *the limiting values* $f(x_{k-1} + 0)$ *and* $f(x_k - 0)$ *should be taken as the values of the function at the end points of* $[x_{k-1}, x_k]$*.

In other words, the domain of any piecewise Hölder function falls into a finite number of closed intervals having no interior points in common on each of which the function belongs to a Hölder class with some positive index.

Each of those intervals will be called *a smoothness section* of the function.

Definition 2. We shall say that a function $f(x)$ is piecewise smooth *on a closed interval* $[a, b]$ *if it is piecewise continuous on* $[a, b]$ *and has a piecewise continuous derivative on* $[a, b]$**, *i.e. if* $f(x)$ *is piecewise continuous on* $[a, b]$ *and its derivative* $f'(x)$ *exists and is continuous everywhere on* $[a, b]$ *except possibly for a finite number of points at each of which the function* $f'(x)$ *has finite right- and left-hand limiting values.*

It is clear that any function piecewise smooth on a closed interval $[a, b]$ is piecewise Hölder on it.

The following *main* theorem holds.

Theorem 10.17. Let a function $f(x)$ *piecewise Hölder on an interval* $[-\pi, \pi]$ *be periodically (with period* 2π*) extended to the whole infinite straight line. Then the Fourier trigonometric series of* $f(x)$ *converges at each point* x *of the infinite line to the value* $f(x) = 1/2\,[f(x - 0) + f(x + 0)]$, *the convergence of that series being uniform on every fixed closed interval inside a smoothness section of* $f(x)$.

Proof. The statement of the theorem about the uniform convergence on every fixed closed interval inside a smoothness section follows immediately from Theorem 10.16. Also following from this theorem is the convergence of the Fourier trigonometric series of $f(x)$ at each *interior* point of a smoothness section of $f(x)$***. It remains to prove the convergence of the Fourier trigonometric series of $f(x)$ at each point of junction of two smoothness sections.

Fix one of such points and denote it by x. Then there are constants M_1 and M_2 such that for any sufficiently small positive t

$$|f(x + t) - f(x + 0)| \leqslant M_1 t^{\alpha_1} \quad (0 < \alpha_1 \leqslant 1), \qquad (10.75)$$

---

* A piecewise Hölder function, as any piecewise continuous function, must have values equal at every point x_k to a half-sum of the right- and left-hand limiting values at that point, i.e. the equation $f(x_k) = 1/2\,[f(x_k - 0) + f(x_k + 0)]$ must hold.

** See Definition 1 in Section 10.4.2.

*** For each interior point of a smoothness section may be taken as a neighbourhood of a closed interval inside that section.

and for any sufficiently small negative t

$$|f(x+t)-f(x-0)|\leqslant M_2\cdot|t|^{\alpha_2}\quad(0<\alpha_2\leqslant 1). \tag{10.76}$$

Denote by M the largest of the numbers M_1 and M_2 and by α the smallest of the numbers α_1 and α_2. Then, with $|t|\leqslant 1$, on the right of each of the inequalities (10.75) and (10.76) we may write $M\cdot|t|^{\alpha}$.

Now fix an arbitrary $\varepsilon>0$ and choose for it $\delta>0$ satisfying (10.70) and so small that both inequalities (10.75) and (10.76) hold for $|t|\leqslant\delta$ and we may take a number $M\cdot|t|^{\alpha}$ on the right-hand sides of those inequalities. Repeating the reasoning of the proof for Theorem 10.15 we arrive at equation (10.71) and to prove the theorem it remains for us to show that the estimates (10.72), (10.73), and (10.74) hold at the fixed point x. In Remark 2 of Section 10.5.5 we have noted that the estimates (10.73) and (10.74) hold for *any function only piecewise continuous and periodical* (*with period* 2π). It remains to prove the validity for all n of the estimate (10.72).

Having in mind that $f(x)=1/2\;[f(x-0)+f(x+0)]$ and that*

$$\int_{-\delta}^{\delta}\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt=2\int_{0}^{\delta}\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt=2\int_{-\delta}^{0}\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt$$

we may rewrite the integral on the left of (10.72) as follows:

$$\frac{1}{\pi}\int_{|t|\leqslant\delta}[f(x+t)-f(x)]\,\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt=$$

$$=\frac{1}{\pi}\int_{0}^{\delta}[f(x+t)-f(x+0)]\,\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt+$$

$$+\frac{1}{\pi}\int_{-\delta}^{0}[f(x+t)-f(x-0)]\,\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt. \tag{10.77}$$

---

* In view of the fact that the function $\varphi(t)=\dfrac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}$ is *even*, i.e. satisfies $\varphi(-t)=\varphi(t)$ for any t. It is easy to see that for such a function $\int_{0}^{\delta}\varphi(t)\,dt=\int_{-\delta}^{0}\varphi(t)\,dt$ (it suffices to make a substitution $t=-\tau$ in one of the integrals) and therefore

$$\int_{-\delta}^{\delta}\varphi(t)\,dt=2\int_{0}^{\delta}\varphi(t)\,dt=2\int_{-\delta}^{0}\varphi(t)\,dt.$$

To evaluate the integrals on the right of (10.77) use inequalities (10.75) and (10.76), writing on the right-hand sides of the inequalities a number $M\,|t|^{\alpha}$. Taking into account the estimate $\dfrac{1}{2\left|\sin\frac{t}{2}\right|}\leqslant\dfrac{\pi}{2\,|t|}$ (with $|t|\leqslant\pi$), already applied in proving Theorem 10.5, and inequality (10.70) we have

$$\left|\frac{1}{\pi}\int\limits_{|t|\leqslant\delta}[f(x+t)-f(x)]\,\frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\,dt\right|\leqslant$$

$$\leqslant\frac{M}{2}\left[\int\limits_{0}^{\delta}t^{\alpha-1}\,dt+\int\limits_{-\delta}^{0}|t|^{\alpha-1}\,dt\right]=\frac{M}{\alpha}\cdot\delta^{\alpha}<\frac{\varepsilon}{3}.$$

Thus the estimate (10.72) and hence the theorem are proved.

Corollary 1. The statement of Theorem 10.17 will clearly be true if we take a function piecewise smooth instead of piecewise Hölder (on $[-\pi,\ \pi]$), *periodically* (*with period* 2π) *extended to the whole infinite straight line.*

We introduce a new concept to formulate one more corollary. Let $0<\alpha\leqslant 1$.

Definition 3. We shall say that a function $f(x)$ satisfies at a given point x **on the right (left) a Hölder condition** *of order* α *if* $f(x)$ *has at the point* x *a right-hand (left-hand) limiting value and if there is a constant* M *such that for all sufficiently small positive (negative)* t

$$\frac{|f(x+t)-f(x+0)|}{t^{\alpha}}\leqslant M\qquad\left(\frac{|f(x+t)-f(x-0)|}{|t|^{\alpha}}\leqslant M\right).$$

It is obvious that if $f(x)$ has at a given point x a right-hand (left-hand) derivative understood to be the limit

$$\lim_{t\to 0+0}\frac{f(x+t)-f(x+0)}{t}\qquad\left(\lim_{t\to 0-0}\frac{f(x+t)-f(x-0)}{t}\right),$$

then $f(x)$ a fortiori satisfies at that point x on the right (left) a Hölder condition of any order $\alpha\leqslant 1$.

Corollary 2 (condition for convergence of a Fourier trigonometric series at a given point). *For the Fourier trigonometric series of a piecewise continuous and periodic function* $f(x)$ (*with period* 2π) *to converge at a given point* x *of an infinite straight line it is sufficient that* $f(x)$ *satisfies a Hölder condition of some positive order* α_1 *at the point* x *on the right and a Hölder condition of some positive order* α_2 *at the point* x *on the left (and it is clearly sufficient that* $f(x)$ *has a right-hand and a left-hand derivative at the point* x).

Proof. It suffices to notice that from the fact that $f(x)$ satisfies a Hölder condition of order α_1 (order α_2) at a point x on the right (left) it follows that there is a constant M_1 (a constant M_2) such that inequality (10.75) (inequality (10.76)) holds for all sufficiently small positive (negative) t. But our proof of Theorem 10.17 uses only inequalities (10.75) and (10.76) and the piecewise continuity and periodicity of $f(x)$.

Example. Without computing the Fourier coefficients of the function

$$f(x)=\begin{cases}\cos x & \text{when} \quad -\pi\leqslant x<0,\\ 1/2 & \text{when} \quad x=0,\\ \sqrt{x} & \text{when} \quad 0<x\leqslant\pi,\end{cases}$$

we may state that the Fourier trigonometric series of that function converges at a point $x = 0$ to a value of $1/2$, for $f(x)$ has at that point a left-hand derivative and satisfies at it on the right a Hölder condition of order $\alpha_2 = 1/2$.

10.5.7. The summability of the Fourier trigonometric series of a continuous function by arithmetic means. We have already noted that the Fourier trigonometric series of an everywhere continuous and periodic function (with period 2π) may be divergent (see Section 10.5.1). We prove that that series is nevertheless always summable (uniformly on the whole infinite line) by the Cesaro method (or the method of arithmetic means)*.

*Theorem 10.18 (Fejér**). If a function $f(x)$ is continuous on an interval $[-\pi, \pi]$ and satisfies the condition $f(-\pi) = f(\pi)$, then the arithmetic means of the partial sums of its Fourier trigonometric series*

$$\sigma_n(x, f)=\frac{S_0(x, f)+S_1(x, f)+\ldots+S_{n-1}(x, f)}{n}$$

converge (to that function) uniformly on $[-\pi, \pi]$ (,and provided a function with period 2π is extended to the whole infinite line, they converge uniformly on the whole infinite line).

Proof. From equation (10.55), for $S_n(x, f)$ we get

$$\sigma_n(x, f)=\frac{1}{n\pi}\int_{-\pi}^{\pi}\frac{f(x+t)}{2\sin\frac{t}{2}}\left[\sum_{k=0}^{n-1}\sin\left(k+\frac{1}{2}\right)t\right]dt. \tag{10.78}$$

To compute the sum in the brackets of (10.78), sum the identity

$$2\sin\frac{t}{2}\sin\left(k+\frac{1}{2}\right)t=\cos kt-\cos(k+1)t$$

over all $k = 0, 1, \ldots, n-1$. As a result we get

$$2\sin\frac{t}{2}\sum_{k=0}^{n-1}\sin\left(k+\frac{1}{2}\right)t=1-\cos nt=2\sin^2\frac{nt}{2}.$$

---

* See Supplement 3 to Chapter 13 of [1].

** L. Fejér (1880-1959) is a Hungarian mathematician. He proved his theorem in 1904.

Using the last equation reduces (10.78) to the form

$$\sigma_n(x, f) = \frac{1}{n\pi} \int_{-\pi}^{\pi} f(x+t) \frac{\sin^2 \frac{nt}{2}}{2\sin^2 \frac{t}{2}}\, dt. \tag{10.79}$$

From (10.79) it immediately follows in turn that

$$\frac{1}{n\pi} \int_{-\pi}^{\pi} \frac{\sin^2 \frac{nt}{2}}{2\sin^2 \frac{t}{2}}\, dt = 1, \tag{10.80}$$

for the left-hand side of (10.80) is equal to the arithmetic mean of the partial sums of the Fourier trigonometric series of the function $f(x) \equiv 1$ and all the partial sums are identically equal to unity (see Section 10.5.2).

Fix an arbitrary $\varepsilon > 0$. According to Theorem 10.9 (Weierstrass) there is a trigonometric polynomial $T(x)$ such that

$$| f(x) - T(x) | \leqslant \varepsilon/2 \tag{10.81}$$

for all x of the infinite line. By the linearity of arithmetic means $\sigma_n(x, f) = \sigma_n(x, f - T) + \sigma_n(x, T)$, so that

$$| \sigma_n(x, f) - T(x) | \leqslant | \sigma_n(x, f - T) | + | \sigma_n(x, T) - T(x) |. \tag{10.82}$$

On writing equation (10.79) for the function $[f(x) - T(x)]$ we get, taking into account the nonnegativeness of the function $\dfrac{\sin^2 \frac{nt}{2}}{2\sin^2 \frac{t}{2}}$ called *the Fejér kernel* and using the estimate (10.81) and equation (10.80),

$$| \sigma_n(x, f-T) | \leqslant \frac{1}{\pi n} \int_{-\pi}^{\pi} | f(x+t) - T(x+t) | \frac{\sin^2 \frac{nt}{2}}{2\sin^2 \frac{t}{2}}\, dt \leqslant$$

$$\leqslant \frac{\varepsilon}{2} \cdot \frac{1}{\pi n} \int_{-\pi}^{\pi} \frac{\sin^2 \frac{nt}{2}}{2\sin^2 \frac{t}{2}}\, dt = \frac{\varepsilon}{2}. \tag{10.83}$$

Inequality (10.83) holds for any n. Now notice that the Fourier trigonometric series of a polynomial $T(x)$ *coincides with that polynomial.* It follows that all partial sums $S_n(x, T)$ beginning with some integer n_0 are equal to $T(x)$. But this allows us to find for the arbitrary $\varepsilon > 0$ fixed above an integer N such that

$$| \sigma_n(x, T) - T(x) | < \varepsilon/2 \tag{10.84}$$

for all $n \geqslant N$ and all x.

From inequalities (10.82), (10.83), and (10.84) we conclude that $| \sigma_n(x, f) - f(x) | < \varepsilon$ for all $n \geqslant N$ and all x. Thus the theorem is proved.

10.5.8. Concluding remarks. **1°.** In solving a number of specific problems one has to expand the function into a Fourier trigonometric series not on the interval $[-\pi, \pi]$ but on $[-l, l]$ where l is an ar-

bitrary positive number. To go over to such a case it suffices to replace the variable x by $\frac{\pi}{l}x$ throughout the above reasoning. Of course, all the established results will remain valid under such a linear change of variable. They will apply to the Fourier trigonometric series

$$\frac{a_0}{2}+\sum_{k=1}^{\infty}\left(a_k\cos\frac{\pi}{l}kx+b_k\sin\frac{\pi}{l}kx\right) \tag{10.85}$$

with the following expressions for the Fourier coefficients

$$\left.\begin{aligned}&a_0=\frac{1}{l}\int_{-l}^{l}f(t)\,dt,\\&a_k=\frac{1}{l}\int_{-l}^{l}f(t)\cos\frac{\pi}{l}kt\,dt,\quad b_k=\frac{1}{l}\int_{-l}^{l}f(t)\sin\frac{\pi}{l}kt\,dt\\&(k=1,\ 2,\ \ldots).\end{aligned}\right\} \tag{10.86}$$

We shall not reformulate any of the established theorems but merely note that in all the statements the interval $[-\pi, \pi]$ should be replaced by $[-l, l]$ and the period 2π by a period $2l$.

2°. Recall that a function $f(x)$ is said to be *even* if it satisfies the condition $f(-x)=f(x)$ and *odd* if it satisfies $f(-x)=-f(x)$.

From the form (10.86) of Fourier trigonometric coefficients it follows that for an even function $f(x)$ all the coefficients b_k ($k=1, 2, \ldots$) are equal to zero and for an odd function $f(x)$ all the coefficients a_k ($k=0, 1, 2, \ldots$) are equal to zero. Thus *an even function $f(x)$ can be expanded into a Fourier trigonometric series only with respect to cosines*

$$\frac{a_0}{2}+\sum_{k=1}^{\infty}a_k\cos\frac{\pi}{l}kx$$

and an odd function $f(x)$ can be expanded into a Fourier trigonometric series only with respect to sines

$$\sum_{k=1}^{\infty}b_k\sin\frac{\pi}{l}kx.$$

3°. We give a very frequently used *complex notation* for the Fourier trigonometric series (10.85).

Using the relations*

$$e^{-i\frac{\pi}{l}kx} = \cos\frac{\pi}{l}kx - i\sin\frac{\pi}{l}kx, \quad e^{i\frac{\pi}{l}kx} = \cos\frac{\pi}{l}kx + i\sin\frac{\pi}{l}kx,$$

it is easy to see that the Fourier trigonometric series (10.85) with the Fourier coefficients (10.86) reduces to the form

$$\sum_{k=-\infty}^{\infty} c_k e^{-i\frac{\pi}{l}kx}, \tag{10.87}$$

where the complex coefficients c_k are of the form

$$c_k = \frac{1}{2l}\int_{-l}^{l} f(t)\, e^{i\frac{\pi}{l}kt}\, dt \tag{10.88}$$

and can be expressed in terms of the coefficients (10.86) by the formulas

$$c_0 = \frac{a_0}{2}, \quad c_{-k} = \frac{a_k - ib_k}{2}, \quad c_k = \frac{a_k + ib_k}{2} \quad (k = 1, 2, \ldots).$$

4°. Especially important for applications is the problem of evaluating a function from approximate Fourier coefficients of that function. A solution of this problem using the so-called *regularization method* is given in the Appendix.

10.6. THE FOURIER INTEGRAL

When a function $f(x)$ is given on the whole infinite line and is not periodic whatever the finite period, it is natural to expand it not into a Fourier trigonometric series, but into the so-called *Fourier integral.*

The present section is devoted to the study of such an expansion.

Throughout it we make the function $f(x)$ obey the requirement of absolute integrability on an infinite line $(-\infty, \infty)$, i.e. we require that there should be an improper integral

$$\int_{-\infty}^{\infty} |f(x)|\, dx. \tag{10.89}$$

Let us agree to use the following terminology.

Definition. We shall say that a function $f(x)$ belongs on an infinite line $(-\infty, \infty)$ to a class L_1 *and write* $f(x) \in L_1(-\infty, \infty)$ *if* $f(x)$ *is integrable (in the proper Riemann sense) on any closed interval and if the improper integral* (10.89) *converges.*

* These relations are immediate consequences of the Euler formula established in Section 1.5.3.

10.6.1. The Fourier transform and its simplest properties.

Lemma 4. If $f(x) \in L_1(-\infty, \infty)$, *then for any point* y *of an infinite line* $-\infty < y < \infty$ *there is an improper integral**

$$\hat{f}(y) = \int_{-\infty}^{\infty} e^{ixy} f(x)\, dx \tag{10.90}$$

called the **Fourier transform** (*or* **image**) *of the function* $f(x)$. *Moreover, the function* $\hat{f}(y)$ *is continuous in* y *at each point of the infinite line and tends to zero as* $|y| \to \infty$, *i.e.*

$$\lim_{|y| \to \infty} |\hat{f}(y)| = 0. \tag{10.91}$$

Proof. It follows from the equation $|e^{ixy} f(x)| = |f(x)|$, from the convergence of the integral (10.89) and from the Weierstrass test (see Theorem 9.7) that the integral (10.90) converges uniformly in y on each closed interval of the infinite line, and hence, by virtue of the continuity of the function e^{ixy} in y, it follows from Theorem 9.9 that the integral (10.90) is continuous in y (on each closed interval, i.e. at each point of the infinite line).

It remains to prove the relation (10.91). Fix an arbitrary $\varepsilon > 0$. In view of the convergence of the integral (10.89) we may fix $A > 0$ such that

$$\int_{-\infty}^{-A} |f(x)|\, dx + \int_{A}^{\infty} |f(x)|\, dx < \frac{\varepsilon}{3}. \tag{10.92}$$

With A thus fixed (by virtue of (10.92)) we have

$$\left| \int_{-\infty}^{\infty} e^{ixy} f(x)\, dx \right| \leqslant \left| \int_{-A}^{A} e^{ixy} f(x)\, dx \right| + \frac{\varepsilon}{3} \tag{10.93}$$

and for the relation (10.91) to be proved it remains for us to establish that the integral on the right of (10.93) is less than $\frac{2}{3}\varepsilon$ for all sufficiently large $|y|$.

Since $f(x)$ is integrable on $[-A, A]$, we can fix a subdivision T of $[-A, A]$ such that for the upper sum S_T of the subdivision**

$$0 < S_T - \int_{-A}^{A} f(x)\, dx < \frac{\varepsilon}{3}. \tag{10.94}$$

* The complex function $\hat{f}(y) = u(y) + iv(y)$ of a real-valued independent variable y is considered as a pair of real-valued functions $u(y)$ and $v(y)$. The continuity of $\hat{f}(y)$ at a given point y is understood to be the continuity at that point of each of the functions $u(y)$ and $v(y)$.

** See Sections 10.2 and 10.3 of [1].

Suppose that the subdivision T is made using the points $-A = x_0 < x_1 < x_2 < \ldots < x_n = A$ and that M_k is the supremum of $f(x)$ on a subinterval $[x_{k-1}, x_k]$ $(k = 1, 2, \ldots, n)$.

We introduce the function

$$\overline{f}_T(x) = \begin{cases} M_k & \text{when } x_{k-1} < x < x_k \quad (k = 1, 2, \ldots, n), \\ 0 & \text{when } x = x_k \quad (k = 0, 1, 2, \ldots, n). \end{cases}$$

Since an integral is independent of the value of the integrand in a finite number of points, it is obvious that

$$\int_{-A}^{A} \overline{f}_T(x)\,dx = \sum_{k=1}^{n} M_k (x_k - x_{k-1}) = S_T,$$

so by (10.94)

$$\int_{-A}^{A} |\overline{f}_T(x) - f(x)|\,dx = \int_{-A}^{A} [\overline{f}_T(x) - f(x)]\,dx < \frac{\varepsilon}{3}. \tag{10.95}$$

Relying on inequality (10.95) and considering that $|e^{ixy}| = 1$ and that $\left|\int_{x_{k-1}}^{x_k} e^{ixy}\,dx\right| \leqslant 2/|y|$ we have

$$\left|\int_{-A}^{A} e^{ixy} f(x)\,dx\right| = \left|\int_{-A}^{A} e^{ixy}[f(x) - \overline{f}_T(x) + \overline{f}_T(x)]\,dx\right| \leqslant$$

$$\leqslant \left|\int_{-A}^{A} e^{ixy}\overline{f}_T(x)\,dx\right| + \left|\int_{-A}^{A} e^{ixy}[\overline{f}_T(x) - f(x)]\,dx\right| \leqslant$$

$$\leqslant \sum_{k=1}^{n} |M_k| \cdot \left|\int_{x_{k-1}}^{x_k} e^{ixy}\,dx\right| + \int_{-A}^{A} |\overline{f}_T(x) - f(x)|\,dx \leqslant$$

$$\leqslant \frac{2}{|y|} \sum_{k=1}^{n} |M_k| + \frac{\varepsilon}{3} \leqslant \frac{2}{3}\varepsilon,$$

provided $|y| > \frac{6}{\varepsilon}\left[\sum_{k=1}^{n} |M_k|\right]$. Thus the lemma is proved.

Corollary. *If $f(x) \in L_1(-\infty, \infty)$, then*

$$\lim_{\lambda\to\infty} \int_{-\infty}^{\infty} \cos\lambda x \cdot f(x)\,dx = 0, \quad \lim_{\lambda\to\infty} \int_{-\infty}^{\infty} \sin\lambda x \cdot f(x)\,dx = 0.$$

10.6.2. Conditions for the expansion of a function into a Fourier integral.

Definition. For each function $f(x)$ of the class $L_1(-\infty, \infty)$ the limit

$$\lim_{\lambda\to\infty}\frac{1}{2\pi}\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy=\lim_{\lambda\to\infty}\frac{1}{2\pi}\int_{-\lambda}^{\lambda}\left[\int_{-\infty}^{\infty} e^{iy(u-x)}f(u)\,du\right]dy$$

(provided it exists) is called the expansion *of the function* into a Fourier integral.

We prove the following *main* theorem.

Theorem 10.19 (condition for the expansion of a function at a given point into a Fourier integral). If $f(x)\in L_1(-\infty, \infty)$ and if $f(x)$ satisfies a Hölder condition of some positive order α_1 $(0<\alpha_1\leqslant 1)$ at a given point x on the right and a Hölder condition of some positive order α_2 $(0<\alpha_2\leqslant 1)$ on the left, then at that point x

$$\lim_{\lambda\to\infty}\frac{1}{2\pi}\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy=\frac{f(x+0)+f(x-0)}{2}. \tag{10.96}$$

Remark 1. At every point x at which the value of $f(x)$ is equal to a half-sum of the right- and left-hand limiting values (in particular, at each point of continuity of $f(x)$) we may write $f(x)$ on the right of (10.96).

Proof of Theorem 10.19. Since the Fourier transform $\hat{f}(y)$ (in view of Lemma 4) is a continuous function of y, for any *positive* λ there is an integral

$$\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy=\int_{-\lambda}^{\lambda} e^{-ixy}\left[\int_{-\infty}^{\infty} e^{iuy}f(u)\,du\right]dy. \tag{10.97}$$

In the integral on the right of (10.97) we may change the order of integration with respect to y and u (since the inner integral converges uniformly with respect to y on any interval $[-\lambda, \lambda]$).

Change the order of integration with respect to y and u, use the equation $e^{iy(u-x)}=\cos y(u-x)+i\sin y(u-x)$,

$$\int_{-\lambda}^{\lambda}\cos y(u-x)\,dy=\frac{\sin\lambda(u-x)}{2(u-x)},\qquad \int_{-\lambda}^{\lambda}\sin y(u-x)\,dy=0$$

and make the substitution $u=x+t$ to get

$$\frac{1}{2\pi}\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\lambda}^{\lambda} e^{iy(u-x)}\,dy\right]f(u)\,du=$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin\lambda(u-x)}{u-x}f(u)\,du=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin\lambda t}{t}f(x+t)\,dt.$$

So for any positive λ

$$\frac{1}{2\pi}\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy=$$

$$=\frac{1}{\pi}\int_{-\infty}^{0}\frac{\sin\lambda t}{t}f(x+t)\,dt+\frac{1}{\pi}\int_{0}^{\infty}\frac{\sin\lambda t}{t}f(x+t)\,dt. \tag{10.98}$$

Now note that for any positive λ we have the equation*

$\int_0^{\infty}\frac{\sin\lambda t}{t}\,dt=\frac{\pi}{2}$ and, therefore, $\int_{-\infty}^{0}\frac{\sin\lambda t}{t}\,dt=\frac{\pi}{2}$.

It follows from the last two equations that for any positive λ

$$\frac{f(x+0)}{2}=\frac{1}{\pi}\int_0^{\infty}f(x+0)\frac{\sin\lambda t}{t}\,dt, \tag{10.99}$$

$$\frac{f(x-0)}{2}=\frac{1}{\pi}\int_{-\infty}^{0}f(x-0)\frac{\sin\lambda t}{t}\,dt. \tag{10.100}$$

Subtracting from (10.98) equations (10.99) and (10.100) we see that for any positive λ

$$\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy-\frac{f(x+0)+f(x-0)}{2}=$$

$$=\frac{1}{\pi}\int_0^{\infty}[f(x+t)-f(x+0)]\frac{\sin\lambda t}{t}\,dt+$$

$$+\frac{1}{\pi}\int_{-\infty}^{0}[f(x+t)-f(x-0)]\frac{\sin\lambda t}{t}\,dt. \tag{10.101}$$

Since $f(x)$ satisfies a Hölder condition of order α_1 on the right and a Hölder condition α_2 on the left, there are constants M_1 and M_2 such that inequality (10.75) holds for all sufficiently small positive t and inequality (10.76) does for all sufficiently small negative t. If we denote by M the largest of the numbers M_1 and M_2 and by α the smallest of the numbers α_1 and α_2, then we may write $M\,|\,t\,|^{\alpha}$ on the right-hand sides of (10.75) and (10.76), these inequalities being true for all positive (respectively, negative) values of t satisfying $|\,t\,|\leqslant\delta$, where δ is an arbitrary, sufficiently small positive number.

* See Section 9.3.

Now we may rewrite relation (10.101) as follows:

$$\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy-\frac{f(x+0)+f(x-0)}{2}=$$

$$=\frac{1}{\pi}\int_{0}^{\delta}[f(x+t)-f(x+0)]\frac{\sin\lambda t}{t}\,dt+$$

$$+\frac{1}{\pi}\int_{-\delta}^{0}[f(x+t)-f(x-0)]\frac{\sin\lambda t}{t}\,dt+$$

$$+\frac{1}{\pi}\int_{|t|\geqslant\delta} f(x+t)\frac{\sin\lambda t}{t}\,dt-$$

$$-\frac{f(x+0)}{\pi}\int_{\delta}^{\infty}\frac{\sin\lambda t}{t}\,dt-\frac{f(x-0)}{\pi}\int_{-\infty}^{-\delta}\frac{\sin\lambda t}{t}\,dt. \qquad (10.102)$$

Fix an arbitrary $\varepsilon>0$ and choose for it $\delta>0$ so small that

$$\frac{M\delta^{\alpha}}{\pi\alpha}<\frac{\varepsilon}{4}. \qquad (10.103)$$

Evaluating the first two integrals on the right of (10.102) with the aid of inequalities (10.75) and (10.76) (with $M\,|\,t\,|^{\alpha}$ on the right-hand sides of them) we have

$$\left|\frac{1}{\pi}\int_{0}^{\delta}[f(x+t)-f(x+0)]\frac{\sin\lambda t}{t}\,dt\right|\leqslant$$

$$\leqslant\frac{1}{\pi}\int_{0}^{\delta}|f(x+t)-f(x+0)|\frac{dt}{t}\leqslant\frac{M}{\pi}\int_{0}^{\delta}t^{\alpha-1}\,dt=\frac{M\delta^{\alpha}}{\pi\alpha}$$

and quite similarly

$$\left|\frac{1}{\pi}\int_{-\delta}^{0}[f(x+t)-f(x-0)]\frac{\sin\lambda t}{t}\,dt\right|\leqslant$$

$$\leqslant\frac{1}{\pi}\int_{-\delta}^{0}|f(x+t)-f(x-0)|\frac{dt}{|t|}\leqslant\frac{M}{\pi}\int_{-\delta}^{0}|t|^{\alpha-1}\,dt=\frac{M\delta^{\alpha}}{\pi\alpha}.$$

From the last two inequalities and from (10.103) we get

$$\left|\frac{1}{\pi}\int_{0}^{\delta}[f(x+t)-f(x+0)]\frac{\sin\lambda t}{t}\,dt\right|+$$

$$+\left|\frac{1}{\pi}\int_{-\delta}^{0}[f(x+t)-f(x-0)]\frac{\sin\lambda t}{t}\,dt\right|<\frac{\varepsilon}{2}. \qquad (10.104)$$

To evaluate the third integral on the right of (10.102) we introduce the function

$$g(t)=\begin{cases}\frac{1}{\pi}\frac{f(x+t)}{t} & \text{when } |t|\geqslant\delta\\ 0 & \text{when } |t|<\delta.\end{cases}$$

Since $g(t)\in L_1(-\infty,\infty)$, by the corollary of Lemma 4

$$\lim_{\lambda\to\infty}\int_{-\infty}^{\infty} g(t)\sin\lambda t\,dt=\lim_{\lambda\to\infty}\frac{1}{\pi}\int_{|t|\geqslant\delta} f(x+t)\frac{\sin\lambda t}{t}\,dt=0,$$

but this means that for the fixed arbitrary $\varepsilon>0$ there is Λ_1 such that

$$\left|\frac{1}{\pi}\int_{|t|\geqslant\delta} f(x+t)\frac{\sin\lambda t}{t}\,dt\right|<\frac{\varepsilon}{4}\ (\text{with } \lambda\geqslant\Lambda_1). \tag{10.105}$$

Finally note that

$$\int_{-\infty}^{-\delta}\frac{\sin\lambda t}{t}\,dt=\int_{\delta}^{\infty}\frac{\sin\lambda t}{t}\,dt=\int_{\lambda\delta}^{\infty}\frac{\sin\tau}{\tau}\,d\tau\to 0$$

as $\lambda\to\infty$. It follows that for the fixed arbitrary $\varepsilon>0$ and the given point x there is Λ_2 such that

$$\left|\frac{f(x+0)}{\pi}\int_{\delta}^{\infty}\frac{\sin\lambda t}{t}\,dt\right|+\left|\frac{f(x-0)}{\pi}\int_{-\infty}^{-\delta}\frac{\sin\lambda t}{t}\,dt\right|<\frac{\varepsilon}{4}$$

$$(\text{with } \lambda\geqslant\Lambda_2). \tag{10.106}$$

Denote by Λ the largest of the numbers Λ_1 and Λ_2. From relations (10.102) and (10.104) to (10.106) we concluded that

$$\left|\int_{-\lambda}^{\lambda} e^{-ixy}\hat{f}(y)\,dy-\frac{f(x+0)+f(x-0)}{2}\right|<\varepsilon\ (\text{with } \lambda\geqslant\Lambda).$$

Thus the theorem is proved.

Corollary. Equation (10.96) *is clearly true if* $f(x)\in L_1(-\infty,\infty)$ *and if* $f(x)$ *has at a given point* x *the right- and left-hand derivatives understood to be the limits of relations* $\lim\limits_{t\to 0+0}\frac{f(x+t)-f(x+0)}{t}$ $\left(\lim\limits_{t\to 0-0}\frac{f(x+t)-f(x-0)}{t}\right)$.

Remark 2. The limit on the left of (10.96) may be written as an improper integral

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixy}\hat{f}(y)\,dy, \tag{10.107}$$

($\hat{f}(y)$ is sometimes called *a Fourier sine transform*),

$$f(x)=\frac{2}{\pi}\int_0^\infty \frac{y\sin yx}{a^2+y^2}\,dy=\begin{cases} e^{-ax} & \text{when } x>0,\\ 0 & \text{when } x=0.\end{cases}$$

10.6.4. Some further properties of the Fourier transformation. We shall discuss here some further properties of Fourier transformation, frequently encountered in applications.

Lemma 5. For some nonnegative integer k, let a function $(1+|x|)^k\cdot f(x)\in L_1(-\infty,\infty)$. Then the Fourier transform (10.90) *of the function $f(x)$ is differentiable k times with respect to the variable y, it being possible to compute the derivative with respect to y of any order m ($m=1, 2, \ldots, k$) by differentiation under the integral* (10.90), *i.e. from the formula*

$$\frac{d^m}{dy^m}\hat{f}(y)=\int_{-\infty}^{\infty} e^{ixy}(ix)^m\cdot f(x)\,dx \quad (m=1, 2, \ldots, k). \tag{10.111}$$

Proof. From the inequality

$$\left|\frac{d^m}{dy^m}e^{ixy}f(x)\right|=|e^{ixy}\cdot(ix)^m f(x)|\leqslant(1+|x|)^k\cdot|f(x)|$$

true for any m ($m=0, 1, \ldots, k$) and from the convergence of the improper integral $\int_{-\infty}^{\infty}(1+|x|)^k\cdot|f(x)|\,dx$ by the Weierstrass test (Theorem 9.7) it follows that the integral on the right of (10.111) converges uniformly in y (on every closed interval) for any $m=0, 1, \ldots, k$. By virtue of Theorem 9.10 this ensures that there is a derivative with respect to y of any order $m=1, 2, \ldots, k$ and that formula (10.111) is true. Thus the lemma is proved.

Lemma 6. Let a function $f(x)$ have at every point x all derivatives up to order $k\geqslant 1$ inclusively, the function $f(x)$ itself and its derivative of order k being absolutely integrable on an infinite line and for any $m=0, 1, \ldots, (k-1)$

$$\lim_{|x|\to\infty}\left[\frac{d^m f(x)}{dx^m}\right]=0. \tag{10.112}$$

Then for the Fourier transform $\hat{f}(y)$ of $f(x)$ as $|y|\to\infty$ we have the estimate

$$|\hat{f}(y)|=o(|y|^{-k}). \tag{10.113}$$

Proof. Consider for any $\lambda>0$ the integral

$$\int_{-\lambda}^{\lambda} e^{ixy}\frac{d^k f(x)}{dx^k}\,dx.$$

Integrating it k times by parts we obtain the formula

$$\int_{-\lambda}^{\lambda} e^{ixy}\frac{d^k f(x)}{dx^k}dx=\left[e^{ixy}\frac{d^{k-1}f(x)}{dx^{k-1}}\right]\Bigg|_{-\lambda}^{\lambda}-\left[iy\cdot e^{ixy}\frac{d^{k-2}f(x)}{dx^{k-2}}\right]\Bigg|_{-\lambda}^{\lambda}+\ldots$$

$$\ldots+(-i)^k\cdot y^k\int_{-\lambda}^{\lambda} e^{ixy}f(x)\,dx.$$

Making in the equation obtained λ approach ∞ and considering that by (10.112) all substitutions vanish we get

$$\int_{-\infty}^{\infty} e^{ixy} \frac{d^k f(x)}{dx^k} dx = (-iy)^k \int_{-\infty}^{\infty} e^{ixy} \cdot f(x)\, dx = (-iy)^k \cdot \hat{f}(y).$$

Considering that by Lemma 4 the integral on the left of the last equation tends to zero as $|y| \to \infty$ we obtain the estimate (10.113). Thus the lemma is proved.

Theorem 10.20. Let a function $f(x)$ and its second derivative be absolutely integrable on an infinite line $(-\infty, \infty)$, the function $f(x)$ itself and its first derivative tending to zero as $|x| \to \infty$. Also let a function $g(x)$ be absolutely integrable on $(-\infty, \infty)$. Then the following equation is true:

$$\int_{-\infty}^{\infty} f(x)\, g(x)\, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(y)\, \hat{g}^*(y)\, dy \tag{10.114}$$

called the generalized Parseval formula or Plancherel* equation. (In (10.114) $\hat{f}(y)$ and $\hat{g}(y)$ are the Fourier transforms of $f(x)$ and $g(x)$ respectively and $\hat{g}^*(y)$ is a complex conjugate of $\hat{g}(y)$.)

Proof. By virtue of Theorem 10.19 at every point x

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixy} \hat{f}(y)\, dy, \tag{10.115}$$

the estimate $|\hat{f}(y)| \leqslant C(1+|y|)^{-2}$ being true by Lemma 6, ensuring that the integral on the right of (10.115) converges absolutely and uniformly (with respect to x) on the whole infinite line.

Multiplying both sides of (10.115) by $g(x)$ and integrating with respect to x between $-\lambda$ and λ we have

$$\int_{-\lambda}^{\lambda} f(x)\, g(x)\, dx = \frac{1}{2\pi} \int_{-\lambda}^{\lambda} g(x) \left[\int_{-\infty}^{\infty} e^{-ixy} \hat{f}(y)\, dy \right] d\,. \tag{10.116}$$

As pointed out above the integral (10.115) converges uniformly in x and so we may change the order of integration with respect to x and y on the right of (10.116) to get

$$\int_{-\lambda}^{\lambda} f(x)\, g(x)\, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\lambda}^{\lambda} e^{ixy} g(x)\, dx \right]^* \hat{f}(y)\, dy \tag{10.117}$$

(* stands for complex conjugation).

By virtue of the inequality

$$\left| \left[\int_{-\lambda}^{\lambda} e^{ixy} g(x)\, dx \right]^* \right| \cdot |\hat{f}(y)| \leqslant \int_{-\infty}^{\infty} |g(x)| \cdot dx \cdot C(1+|y|)^{-2}$$

and the Weierstrass test the integral on the right of (10.117) converges uniformly with respect to λ on the infinite line $-\infty < \lambda < \infty$. In (10.117) therefore we can proceed to the limit as $\lambda \to \infty$, while proceeding on the right of (10.117) to the limit under the integral. Thus the theorem is proved.

* M. Plancherel (b. 1885) is a French mathematician.

10.7. MULTIPLE FOURIER TRIGONOMETRIC SERIES AND INTEGRALS

10.7.1. A multiple Fourier trigonometric series and its rectangular and spherical partial sums. Let a function of N variables $f(x_1, x_2, \ldots, x_N)$ be defined and integrable in an N-dimensional cube $-\pi \leqslant x_k \leqslant \pi$ $(k = 1, 2, \ldots, N)$. We shall denote that cube by Π. It is convenient to write the multiple trigonometric series of such a function straight in complex form, using the concept of a scalar product of two N-dimensional vectors to abridge the notation.

Let $x = (x_1, x_2, \ldots, x_N)$ be a vector with arbitrary real-valued coordinates $x_1, x_2, \ldots, x_N$, and let $n = (n_1, n_2, \ldots, n_N)$ be a vector with *integral* coordinates $n_1, n_2, \ldots, n_N$.

The multiple Fourier trigonometric series of a function $f(x) = f(x_1, x_2, \ldots, x_N)$ is a series of the form

$$\sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_N=-\infty}^{\infty} \hat{f}_n \cdot e^{-i(xn)}, \tag{10.118}$$

where the numbers $\hat{f}_n$ called *Fourier coefficients* are defined by the equations

$$\hat{f}_n = \hat{f}_{n_1 n_2 \ldots n_N} = (2\pi)^{-N} \int \cdots \int_{\Pi} f(y_1, \ldots, y_N) \times$$
$$\times e^{i(y_1 n_1 + \ldots + y_N n_N)} \, dy_1 \ldots dy_N, \tag{10.119}$$

and (xn) denotes a scalar product of vectors x and n equal to $x_1 n_1 + \ldots + x_N n_N$.

Of course a multiple Fourier trigonometric series (10.118) may be regarded as a Fourier series with respect to an orthonormal system (in an N-dimensional cube Π)* formed using all possible products of elements of the one-dimensional trigonometric system in variables $x_1, x_2, \ldots, x_N$ respectively. It is customary to call that orthonormal system *a multiple trigonometric system*.

For this, as for any orthonormal system, there is *a Bessel inequality* which is of the form

$$\sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_N=-\infty}^{\infty} |\hat{f}_n|^2 \leqslant (2\pi)^{-N} \int \cdots \int_{\Pi} f^2(x_1, \ldots, x_N) \, dx_1 \ldots dx_N, \tag{10.120}$$

where $f(x_1, \ldots, x_N)$ is any function continuous in an N-dimensional cube Π.

* A scalar product of any two functions is defined to be an integral over Π of the product of those functions.

Consider the problem of convergence of a multiple Fourier trigonometric series. If that series *does not converge* at a given point $x = (x_1, \ldots, x_N)$ *absolutely*, then its convergence (by Theorem 13.10 (Riemann) of [11]) depends on the sequence of its terms (or equivalently on the order of summation over the indices $n_1, n_2, \ldots, n_N$).

Of wide use are two methods of summing a multiple Fourier trigonometric series, *the spherical and the rectangular one.*

The spherical partial sums of a multiple Fourier trigonometric series (10.118) are sums of the form

$$S_\lambda(x, f) = \sum_{|n| \leqslant \lambda} \hat{f}_n e^{-i(x n)}$$

taken over all integral values $n_1, n_2, \ldots, n_N$ satisfying the condition $|n| = \sqrt{n_1^2 + n_2^2 + \ldots n_N^2} \leqslant \lambda$.

A multiple Fourier trigonometric series (10.118) *is said to be summable at a given point x by the spherical method if there is a limit* $\lim\limits_{\lambda \to \infty} S\,\Lambda(x, f)$ *at that point.*

The rectangular partial sums of a multiple Fourier trigonometric series (10.118) are sums of the form

$$S_{m_1 m_2 \ldots m_N}(x, f) = \sum_{n_1 = -m_1}^{m_1} \ldots \sum_{n_N = -m_N}^{m_N} \hat{f}_n e^{-i(x n)}.$$

A multiple Fourier trigonometric series (10.118) *is said to be summable at a given point x by the rectangular (or Princegeime) method if there is a limit*

$$\lim_{\substack{m_1 \to \infty \\ m_2 \to \infty \\ \cdots\cdots \\ m_N \to \infty}} S_{m_1 m_2 \ldots m_N}(x, f)$$

at that point (as each of the indices $m_1, m_2, \ldots, m_N$ tends to infinity).

The summation methods both have their advantages and drawbacks. In considering a multiple Fourier trigonometric series with respect to an orthonormal system it is natural to arrange its terms in the order of increasing $|n|$ and deal with spherical partial sums.

Rectangular partial sums are used in investigating the behaviour of multiple power series near the boundary of the domain of convergence. It should be noted that the definition of the sum of a series as the limit of rectangular sums (in contrast to the definition based on the limit of spherical sums) imposes no constraints on the infinite set of partial sums of that kind.

Before formulating the conditions for the convergence of a multiple Fourier trigonometric series we define some smoothness characteristics of a function of N variables.

10.7.2. The modulus of continuity and Hölder classes for a function of N variables. Let a function $f(x) = f(x_1, x_2, \ldots, x_N)$ be defined and continuous in an N-dimensional domain D.

Definition 1. For every $\delta > 0$, **the modulus of continuity** *of a function $f(x)$ in a domain D is the supremum of the absolute value of the difference $|f(x') - f(x'')|$ on the set of all points x' and x'' in D whose distance $\rho(x', x'')$ from each other is less than δ.*

We shall denote the modulus of continuity of $f(x)$ in D by $\omega(\delta, f)$.

Definition 2. For any $\varkappa$ of the half-open interval $0 < \varkappa \leqslant 1$, we shall say that a function $f(x)$ belongs in a domain D to a Hölder class $C^{\varkappa}$ *with an index $\varkappa$ and write $f(x) \in C^{\varkappa}(D)$ if the modulus of continuity of $f(x)$ in D is of order $\omega(\delta, f) = o(\delta^{\varkappa})$ when $0 < \varkappa < 1$ and $\omega(\delta, f) = O(\delta)$ when $\varkappa = 1$.*

Now let α be *any* (not necessarily whole) *positive* number. We can always represent it as $\alpha = r + \varkappa$, where r is an integer and $\varkappa$ is in the half-open interval $0 < \varkappa \leqslant 1$.

Definition 3. We shall say that a function $f(x)$ belongs in a domain D to a Hölder class C^{α} *with an index $\alpha > 0$ and write $f(x) \in C^{\alpha}(D)$ if all partial derivatives of $f(x)$ of order r are continuous in D and every partial derivative of order r belongs to the class $C^{\varkappa}(D)$ introduced in Definition 2.*

10.7.3. Conditions for convergence of a multiple Fourier trigonometric series. We first establish the simplest conditions for absolute and uniform convergence of a multiple Fourier trigonometric series.

Theorem 10.21. If a function $f(x)$ is periodically (with period 2π in each of the variables) extended to the whole of a space E^N and has in E^N continuous derivatives of order $s = [N/2] + 1$, where $[N/2]$ is the integral part of the number $N/2$, then the multiple Fourier trigonometric series of $f(x)$ converges (to that function) absolutely and uniformly in the whole of E^N.

Proof. Let us agree to denote by $\left(\widehat{\dfrac{\partial^m f}{\partial x_k^m}}\right)_n$ the Fourier coefficient of a derivative $\dfrac{\partial^m f}{\partial x_k^m}$ with index $n = (n_1, \ldots, n_N)$. Integrating by parts we get $\left(\widehat{\dfrac{\partial f}{\partial x_k}}\right)_n = i n_k \hat{f}_n$ (for any $k = 1, 2, \ldots, N$), so that $\left|\sum\limits_{k=1}^{N} \left(\widehat{\dfrac{\partial f}{\partial x_k}}\right)_n\right| = |\hat{f}_n| \, (|n_1| + \ldots + |n_N|)$ and therefore

$$|\hat{f}_n| = (|n_1| + \ldots + |n_N|)^{-1} \sum_{k=1}^{N} \left|\left(\widehat{\frac{\partial f}{\partial x_k}}\right)_n\right|. \tag{10.121}$$

Formula (10.121) is true not only for the function f but also for every partial derivative of f up to the order $(s - 1)$ inclusively.

This immediately yields the relation

$$|\hat{f}_n| \leqslant (|n_1|+\ldots+|n_N|)^{-s} \sum_{s_1+\ldots+s_N=s} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right| \tag{10.122}$$

the sum on the right of which is taken over all nonnegative integers $s_1, s_2, \ldots, s_N$ satisfying $s_1+s_2+\ldots+s_N=s$ (so that the number of terms in the sum is equal to N^s). From (10.122) it follows in turn that*

$$|\hat{f}_n| \leqslant \frac{1}{2}(|n_1|+\ldots+|n_N|)^{-2s}+$$

$$+\frac{N^s}{2} \sum_{s_1+\ldots+s_N=s} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right|^2.$$

Considering that $s=\frac{N}{2}+\varepsilon$, where $\varepsilon=1$ for an even N and $\varepsilon=\frac{1}{2}$ for an odd N and that

$$(|n_1|+\ldots+|n_N|)^{-2s}=(|n_1|+\ldots+|n_N|)^{-N-2\varepsilon} \leqslant$$

$$\leqslant |n_1|^{-1-\frac{2\varepsilon}{N}} \ldots |n_N|^{-1-\frac{2\varepsilon}{N}},$$

we get from (10.123)

$$|\hat{f}_n| \leqslant \frac{1}{2}|n_1|^{-1-\frac{2\varepsilon}{N}} \ldots |n_N|^{-1-\frac{2\varepsilon}{N}}+$$

$$+\frac{N^s}{2} \cdot \sum_{s_1+\ldots+s_N=s} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right|^2. \tag{10.124}$$

For the absolute and uniform convergence of a multiple Fourier trigonometric series (10.118) it suffices (by virtue of the Weierstrass test) to prove the convergence of the number series

$$\sum_{n_1=-\infty}^{\infty} \ldots \sum_{n_N=-\infty}^{\infty} |\hat{f}_n|$$

majorizing it, but (by inequality (10.124)) the convergence of this last series is a direct consequence of the convergence of the num-

* We use the inequalities $|a|\cdot|b| \leqslant \frac{a^2}{2}+\frac{b^2}{2}$ and $(|a_1|+\ldots+|a_p|)^2 \leqslant \leqslant p(a_1^2+\ldots+a_p^2)$.

10.7.2. The modulus of continuity and Hölder classes for a function of N variables. Let a function $f(x) = f(x_1, x_2, \ldots, x_N)$ be defined and continuous in an N-dimensional domain D.

Definition 1. *For every* $\delta > 0$, **the modulus of continuity** *of a function* $f(x)$ *in a domain* D *is the supremum of the absolute value of the difference* $|f(x') - f(x'')|$ *on the set of all points* x' *and* x'' *in* D *whose distance* $\rho(x', x'')$ *from each other is less than* δ.

We shall denote the modulus of continuity of $f(x)$ in D by $\omega(\delta, f)$.

Definition 2. *For any* $\varkappa$ *of the half-open interval* $0 < \varkappa \leqslant 1$, *we shall say that a function* $f(x)$ *belongs in a domain* D *to a* Hölder class $C^\varkappa$ *with an index* $\varkappa$ *and write* $f(x) \in C^\varkappa(D)$ *if the modulus of continuity of* $f(x)$ *in* D *is of order* $\omega(\delta, f) = o(\delta^\varkappa)$ *when* $0 < \varkappa < 1$ *and* $\omega(\delta, f) = O(\delta)$ *when* $\varkappa = 1$.

Now let α be *any* (not necessarily whole) *positive* number. We can always represent it as $\alpha = r + \varkappa$, where r is an integer and $\varkappa$ is in the half-open interval $0 < \varkappa \leqslant 1$.

Definition 3. *We shall say that a function* $f(x)$ *belongs in a domain* D *to a* Hölder class C^α *with an index* $\alpha > 0$ *and write* $f(x) \in C^\alpha(D)$ *if all partial derivatives of* $f(x)$ *of order* r *are continuous in* D *and every partial derivative of order* r *belongs to the class* $C^\varkappa(D)$ *introduced in Definition 2.*

10.7.3. Conditions for convergence of a multiple Fourier trigonometric series. We first establish the simplest conditions for absolute and uniform convergence of a multiple Fourier trigonometric series.

Theorem 10.21. *If a function* $f(x)$ *is periodically (with period* 2π *in each of the variables) extended to the whole of a space* E^N *and has in* E^N *continuous derivatives of order* $s = [N/2] + 1$, *where* $[N/2]$ *is the integral part of the number* $N/2$, *then the multiple Fourier trigonometric series of* $f(x)$ *converges (to that function) absolutely and uniformly in the whole of* E^N.

Proof. Let us agree to denote by $\left(\widehat{\dfrac{\partial^m f}{\partial x_k^m}}\right)_n$ the Fourier coefficient of a derivative $\dfrac{\partial^m f}{\partial x_k^m}$ with index $n = (n_1, \ldots, n_N)$. Integrating by parts we get $\left(\widehat{\dfrac{\partial f}{\partial x_k}}\right)_n = i n_k \hat{f}_n$ (for any $k = 1, 2, \ldots, N$), so that $\left|\sum\limits_{k=1}^{N} \left(\widehat{\dfrac{\partial f}{\partial x_k}}\right)_n\right| = |\hat{f}_n| \, (|n_1| + \ldots + |n_N|)$ and therefore

$$|\hat{f}_n| = (|n_1| + \ldots + |n_N|)^{-1} \sum_{k=1}^{N} \left|\left(\widehat{\frac{\partial f}{\partial x_k}}\right)_n\right|. \tag{10.121}$$

Formula (10.121) is true not only for the function f but also for every partial derivative of f up to the order $(s - 1)$ inclusively.

This immediately yields the relation

$$|\hat{f}_n| \leqslant (|n_1| + \ldots + |n_N|)^{-s} \sum_{s_1 + \ldots + s_N = s} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right| \tag{10.122}$$

the sum on the right of which is taken over all nonnegative integers $s_1, s_2, \ldots, s_N$ satisfying $s_1 + s_2 + \ldots + s_N = s$ (so that the number of terms in the sum is equal to N^s). From (10.122) it follows in turn that*

$$|\hat{f}_n| \leqslant \frac{1}{2} (|n_1| + \ldots + |n_N|)^{-2s} +$$

$$+ \frac{N^s}{2} \sum_{s_1 + \ldots + s_N = s} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right|^2.$$

Considering that $s = \frac{N}{2} + \varepsilon$, where $\varepsilon = 1$ for an even N and $\varepsilon = \frac{1}{2}$ for an odd N and that

$$(|n_1| + \ldots + |n_N|)^{-2s} = (|n_1| + \ldots + |n_N|)^{-N-2\varepsilon} \leqslant$$

$$\leqslant |n_1|^{-1-\frac{2\varepsilon}{N}} \ldots |n_N|^{-1-\frac{2\varepsilon}{N}},$$

we get from (10.123)

$$|\hat{f}_n| \leqslant \frac{1}{2} |n_1|^{-1-\frac{2\varepsilon}{N}} \ldots |n_N|^{-1-\frac{2\varepsilon}{N}} +$$

$$+ \frac{N^s}{2} \cdot \sum_{s_1 + \ldots + s_N = s} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right|^2. \tag{10.124}$$

For the absolute and uniform convergence of a multiple Fourier trigonometric series (10.118) it suffices (by virtue of the Weierstrass test) to prove the convergence of the number series

$$\sum_{n_1=-\infty}^{\infty} \ldots \sum_{n_N=-\infty}^{\infty} |\hat{f}_n|$$

majorizing it, but (by inequality (10.124)) the convergence of this last series is a direct consequence of the convergence of the num-

* We use the inequalities $|a| \cdot |b| \leqslant \frac{a^2}{2} + \frac{b^2}{2}$ and $(|a_1| + \ldots + |a_p|)^2 \leqslant \leqslant p(a_1^2 + \ldots + a_p^2)$.

ber series $\sum_{n_k=-\infty}^{\infty} |n_k|^{-1-\frac{2r}{N}}$ for any k and the convergence of the series

$$\sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_N=-\infty}^{\infty} \left| \left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}} \right)_n \right|^2$$

for any $s_1, s_2, \ldots, s_N$ that follows from the Bessel inequality (10.120) for the continuous function $\dfrac{\partial^s f}{\partial x_1^{s_1} \ldots \partial x_N^{s_N}}$.

The fact that it is to the function $f(x)$ that the multiple Fourier trigonometric series (10.118) converges follows from the completeness of a multiple trigonometric system*. Indeed, if the series (10.118) uniformly converged to some function $g(x)$, then from the possibility of term-by-term integration of such a series it would follow that all Fourier coefficients of $g(x)$ coincided with the corresponding Fourier coefficients of $f(x)$. But then the difference $[f(x) - g(x)]$ would be orthogonal to all the elements of the multiple trigonometric system and (by the completeness of the system) equal to zero. Thus the theorem is proved.

Remark 1. Theorem 10.21 may be revised. The following *statement* is true**: *if a function $f(x)$ is periodic in each of the variables (with period 2π) and belongs in a space E^N to a Hölder class C^α, with $\alpha > N/2$, then the multiple Fourier trigonometric series of $f(x)$ converges (to that function) absolutely and uniformly in the whole space E^N.*

Determining conditions for *nonabsolute* convergence of a multiple trigonometric series requires a finer technique.

We shall state without proof the conditions for the summability of a multiple Fourier trigonometric series by the spherical and the rectangular method.

*Theorem 10.22. If a function $f(x_1, \ldots, x_N)$ of $N \geqslant 2$ variables is periodic in each of the variables (with period 2π) and belongs in a space E^N to a Hölder class C^α, with $\alpha \geqslant (N-1)/2$, then the spherical partial sums of the multiple Fourier trigonometric series of $f(x_1, \ldots, x_N)$ converge to that function uniformly in the whole space E^N***.*

---

* The completeness of the multiple trigonometric system follows immediately from the completeness of the constituent one-dimensional trigonometric systems the product of which it is.

** This statement is very simply obtained from Lemma 3.1 proved in the paper by V.A. Ilyin and Sh.A. Alimov. "Conditions for the convergence of spectral decompositions that correspond to self-adjoint extensions of elliptic operators, I" (*Differentsialnye Uravneniya*, Vol. 7, No. 4 (1971), pp. 670-710).

*** This theorem follows from the more general statements proved in the paper "Problems of localization and convergence of Fourier series in fundamental systems of functions of the Laplacian operator" by V.A. Ilyin (*Uspekhi Matematicheskikh Nauk*, Vol. 23, 2 (1968), pp. 61-120) and in the paper cited in the preceding footnote.

*Theorem 10.23. For any positive α less than $(N-1)/2$ and any point x_0 of an N-dimensional cube Π there is a function $f(x_1, x_2, \ldots, x_N)$ of $N \geqslant 2$ variables periodic in each of the variables (with period 2π), belonging in E^N to a class C^α, vanishing in some δ-neighbourhood of x_0 and such that the spherical partial sums of the multiple Fourier trigonometric series of that function have no limit at x_0**.

Theorems 10.22 and 10.23 establish *final* conditions (in Hölder classes C^α) for the convergence of the spherical partial sums of a periodic function $f(x_1, \ldots, x_N)$. According to those theorems there is uniform convergence of spherical partial sums when $\alpha \geqslant (N-1)/2$, but when $\alpha < (N-1)/2$ even the principle of localization fails to hold for them (however smooth a function f may be in a neighbourhood of x_0, its belonging to a class $C^\alpha(E^N)$ when $\alpha < (N-1)/2$ does not ensure that its spherical partial sums converge at x_0).

Final conditions (in Hölder classes C^α) for the convergence of the rectangular partial sums of a multiple Fourier trigonometric series have been established by L.V. Zhizhiashvili**.

Theorem 10.24. If a function $f(x_1, \ldots, x_N)$ of N variables is periodic in each of them (with period 2π) and belongs in E^N to a class C^α, with any $\alpha > 0$, then the rectangular partial sums of the multiple Fourier trigonometric series of $f(x_1, \ldots, x_N)$ converge (to that function) uniformly in E^N.

Remark 2. Note that as far back as 1928 L. Tonelli*** established that the continuity alone of a function $f(x_1, \ldots, x_N)$ of $N \geqslant 2$ variables fails to ensure not only the uniform convergence but also the principle of localization of the rectangular partial sums of the multiple Fourier trigonometric series of the function (there is a function periodic in each of the variables (with period 2π), continuous in E^N, vanishing in some δ-neighbourhood of a given point x_0 and such that the rectangular partial sums of that function diverge at x_0).

10.7.4. On the expansion of a function into an N-fold multiple Fourier integral. Let a function of $N \geqslant 2$ variables, $f(x_1, \ldots, x_N) = f(x)$, allow the existence of an improper integral

$$\int_{E^N} \cdots \int f(x_1, \ldots, x_N)\, dx_1 \ldots dx_N. \tag{10.125}$$

The Fourier transform of such a function is the quantity

$$\hat{f}(y_1, \ldots, y_N) = \hat{f}(y) = \int_{E^N} \cdots \int e^{i(xy)} f(x_1, \ldots, x_N)\, dx_1 \ldots dx_N.$$

In close analogy with Lemma 4 it can be proved that $\hat{f}(y)$ is a continuous function of y everywhere in E^N and tends to zero as $|y| = \sqrt{y_1^2 + \ldots + y_N^2} \to \infty$.

The limit

$$\lim_{\lambda\to\infty} \int_{|y|\leqslant\lambda} \cdots \int \hat{f}(y_1, \ldots, y_N)\, e^{-i(x,y)}\, dy_1 \ldots dy_N \tag{10.126}$$

* This theorem is a special case of a more general statement proved in Chapter 3 of the paper by V.A. Ilyin cited in the preceding footnote.

** "On Conjugate Functions and Trigonometric Series" by L.V. Zhizhiashvili. Thesis for a Doctor's degree. Moscow: MGU (1967).

*** L. Tonelli (1885-1946) is an Italian mathematician.

ber series $\sum_{n_k=-\infty}^{\infty} |n_k|^{-1-\frac{2r}{N}}$ for any k and the convergence of the series

$$\sum_{n_1=-\infty}^{\infty} \cdots \sum_{n_N=-\infty}^{\infty} \left|\left(\frac{\widehat{\partial^s f}}{\partial x_1^{s_1} \dots \partial x_N^{s_N}}\right)_n\right|^2$$

for any $s_1, s_2, \dots, s_N$ that follows from the Bessel inequality (10.120) for the continuous function $\frac{\partial^s f}{\partial x_1^{s_1} \dots \partial x_N^{s_N}}$.

The fact that it is to the function $f(x)$ that the multiple Fourier trigonometric series (10.118) converges follows from the completeness of a multiple trigonometric system*. Indeed, if the series (10.118) uniformly converged to some function $g(x)$, then from the possibility of term-by-term integration of such a series it would follow that all Fourier coefficients of $g(x)$ coincided with the corresponding Fourier coefficients of $f(x)$. But then the difference $[f(x) - g(x)]$ would be orthogonal to all the elements of the multiple trigonometric system and (by the completeness of the system) equal to zero. Thus the theorem is proved.

Remark 1. Theorem 10.21 may be revised. The following *statement* is true**: *if a function $f(x)$ is periodic in each of the variables (with period 2π) and belongs in a space E^N to a Hölder class C^α, with $\alpha > N/2$, then the multiple Fourier trigonometric series of $f(x)$ converges (to that function) absolutely and uniformly in the whole space E^N.*

Determining conditions for *nonabsolute* convergence of a multiple trigonometric series requires a finer technique.

We shall state without proof the conditions for the summability of a multiple Fourier trigonometric series by the spherical and the rectangular method.

Theorem 10.22. *If a function $f(x_1, \dots, x_N)$ of $N \geqslant 2$ variables is periodic in each of the variables (with period 2π) and belongs in a space E^N to a Hölder class C^α, with $\alpha \geqslant (N-1)/2$, then the spherical partial sums of the multiple Fourier trigonometric series of $f(x_1, \dots, x_N)$ converge to that function uniformly in the whole space E^N***.*

---

* The completeness of the multiple trigonometric system follows immediately from the completeness of the constituent one-dimensional trigonometric systems the product of which it is.

** This statement is very simply obtained from Lemma 3.1 proved in the paper by V.A. Ilyin and Sh.A. Alimov. "Conditions for the convergence of spectral decompositions that correspond to self-adjoint extensions of elliptic operators, I" (*Differentsialnye Uravneniya*, Vol. 7, No. 4 (1971), pp. 670-710).

*** This theorem follows from the more general statements proved in the paper "Problems of localization and convergence of Fourier series in fundamental systems of functions of the Laplacian operator" by V.A. Ilyin (*Uspekhi Matematicheskikh Nauk*, Vol. 23, 2 (1968), pp. 61-120) and in the paper cited in the preceding footnote.

Theorem 10.23. For any positive α less than $(N-1)/2$ and any point x_0 of an N-dimensional cube Π there is a function $f(x_1, x_2, \ldots, x_N)$ of $N \geqslant 2$ variables periodic in each of the variables (with period 2π), belonging in E^N to a class C^α, vanishing in some δ-neighbourhood of x_0 and such that the spherical partial sums of the multiple Fourier trigonometric series of that function have no limit at x_0.*

Theorems 10.22 and 10.23 establish *final* conditions (in Hölder classes C^α) for the convergence of the spherical partial sums of a periodic function $f(x_1, \ldots, x_N)$. According to those theorems there is uniform convergence of spherical partial sums when $\alpha \geqslant (N-1)/2$, but when $\alpha < (N-1)/2$ even the principle of localization fails to hold for them (however smooth a function f may be in a neighbourhood of x_0, its belonging to a class $C^\alpha(E^N)$ when $\alpha < (N-1)/2$ does not ensure that its spherical partial sums converge at x_0).

Final conditions (in Hölder classes C^α) for the convergence of the rectangular partial sums of a multiple Fourier trigonometric series have been established by L.V. Zhizhiashvili**.

Theorem 10.24. If a function $f(x_1, \ldots, x_N)$ of N variables is periodic in each of them (with period 2π) and belongs in E^N to a class C^α, with any $\alpha > 0$, then the rectangular partial sums of the multiple Fourier trigonometric series of $f(x_1, \ldots, x_N)$ converge (to that function) uniformly in E^N.

Remark 2. Note that as far back as 1928 L. Tonelli*** established that the continuity alone of a function $f(x_1, \ldots, x_N)$ of $N \geqslant 2$ variables fails to ensure not only the uniform convergence but also the principle of localization of the rectangular partial sums of the multiple Fourier trigonometric series of the function (there is a function periodic in each of the variables (with period 2π), continuous in E^N, vanishing in some δ-neighbourhood of a given point x_0 and such that the rectangular partial sums of that function diverge at x_0).

10.7.4. On the expansion of a function into an N-fold multiple Fourier integral. Let a function of $N \geqslant 2$ variables, $f(x_1, \ldots, x_N) = f(x)$, allow the existence of an improper integral

$$\int \cdots \int_{E^N} f(x_1, \ldots, x_N)\, dx_1 \ldots dx_N. \tag{10.125}$$

The Fourier transform of such a function is the quantity

$$\hat{f}(y_1, \ldots, y_N) = \hat{f}(y) = \int \cdots \int_{E^N} e^{i(xy)} f(x_1, \ldots, x_N)\, dx_1 \ldots dx_N.$$

In close analogy with Lemma 4 it can be proved that $\hat{f}(y)$ is a continuous function of y everywhere in E^N and tends to zero as $|y| = \sqrt{y_1^2 + \ldots + y_N^2} \to \infty$.

The limit

$$\lim_{\lambda\to\infty} \int \cdots \int_{|y|\leqslant\lambda} \hat{f}(y_1, \ldots, y_N)\, e^{-i(x,y)}\, dy_1 \ldots dy_N \tag{10.126}$$

* This theorem is a special case of a more general statement proved in Chapter 3 of the paper by V.A. Ilyin cited in the preceding footnote.

** "On Conjugate Functions and Trigonometric Series" by L.V. Zhizhiashvili. Thesis for a Doctor's degree. Moscow: MGU (1967).

*** L. Tonelli (1885-1946) is an Italian mathematician.

(provided that this limit exists) is called *the expansion of the function $f_l(x)$ into an N-fold multiple Fourier integral.*

The following two statements are true*.

1°. If a function of $N \geqslant 2$ variables, $f(x_1, \ldots, x_N)$, vanishes outside some bounded domain and belongs in the whole of a space E^N to a Hölder class C^α, with $\alpha \geqslant (N-1)/2$, then the expansion of that function into an N-fold multiple Fourier integral (10.126) converges (to that function) uniformly in the whole space E^N.

2°. For any positive α less than $(N-1)/2$ and any point x_0 there is a function of $N \geqslant 2$ variables, $f(x_1, \ldots, x_N)$, different from zero only in a bounded domain, belonging in E^N to a class C^α, vanishing in some δ-neighbourhood of x_0 and such that for that function the limit (10.126) does not exist at the point x_0.

Statements 1° and 2° establish final conditions (in Hölder classes C^α) for the convergence of the expansion into an N-fold multiple Fourier integral of any function equal to zero outside some bounded domain of a space E^N. According to those statements there is a uniform convergence (in any bounded domain) of the expansion into an N-fold multiple Fourier integral when $\alpha \geqslant (N-1)/2$, but when $\alpha < (N-1)/2$ even the principle of localization fails to hold for the expansion into an N-fold multiple Fourier integral (however smooth a function f may be in a neighbourhood of a point x_0, its belonging in the whole of E^N to a class C^α, with $\alpha < (N-1)/2$, does not ensure that the expansion of that function into an N-fold multiple Fourier integral converges at x_0).

---

* Both statements follow from the more general ones proved in the paper by Sh.A. Alimov and V.A. Ilyin "Conditions for the convergence of spectral decompositions that correspond to self-adjoint extensions of elliptic operators, II" (*Differentsialnye Uravneniya*, Vol. 7, No. 5 (1971), pp. 851-882).

CHAPTER 11

HILBERT SPACE

In this chapter we study an important subclass of infinite-dimensional Euclidean spaces, the so-called *Hilbert spaces*.

We establish a special representation, important for applications, of any linear function of the elements of such a space (this function is generally called a *linear functional*), we also establish that it is possible to choose a subsequence, convergent in some weaker sense, of any infinite set of elements of a Hilbert space, bounded in the norm (this property is termed a *weak compactness* of a ball in a Hilbert space).

Particular attention is given to the study of orthonormal systems of elements of a Hilbert space. We establish the equivalence for such systems of the concepts of closure and completeness introduced in Section 10.2. and prove the famous Riesz-Fisher theorem according to which any sequence of numbers, the series of whose squares converges, is a sequence of Fourier coefficients of some element of a Hilbert space in an expansion with respect to a preassigned orthonormal system of elements of that space. In the last section we prove that the so-called *completely continuous self-adjoint operators* acting in Hilbert space have eigenvalues.

11.1. THE SPACE l^2

11.1.1. The space l^2. Consider a set whose elements are all possible sequences of real numbers $(x_1, x_2, \ldots, x_n, \ldots)$ such that the series made up of the squares of those numbers

$$\sum_{k=1}^{\infty} x_k^2 \tag{11.1}$$

is convergent. The elements of such a set will be denoted (as vectors) by semi-bold Latin letters: $\mathbf{x} = (x_1, x_2, \ldots, x_n, \ldots)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n, \ldots)$, and so on. The numbers $x_1, x_2, \ldots, x_n, \ldots$ will be called the coordinates of an element $\mathbf{x} = (x_1, x_2, \ldots, x_n, \ldots)$.

We define the operations of addition of elements and of multiplication of elements by real numbers. A *sum* of two elements $\mathbf{x} =$

$= (x_1, x_2, \ldots, x_n, \ldots)$ and $y = (y_1, y_2, \ldots, y_n, \ldots)$ is an element $z = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n, \ldots)$*. It will be designated $z = x + y$. *A product of an element* $x = (x_1, x_2, \ldots, x_n, \ldots)$ *by a real number* λ is an element designated λx or $x\lambda$ and equal to $(\lambda x_1, \lambda x_2, \ldots, \lambda x_n, \ldots)$. It is easy to verify that the set we have defined is a *linear space*, i.e. to verify that all the axioms relating to the addition of elements and to the multiplication of elements by real numbers are true**.

We now introduce in that set a scalar product of any two elements $x = (x_1, x_2, \ldots, x_n, \ldots)$ and $y = (y_1, y_2, \ldots y_n, \ldots)$, defining it to be the sum of a series***

$$\sum_{k=1}^{\infty} x_k y_k.$$

So we set $(x, y) = \sum_{k=1}^{\infty} x_k y_k$. It is easy to verify that all the four axioms of a scalar product are true. (They can be found in Section 10.1, and verification of their validity for the space under study will be left to the reader.)

Thus the set we have introduced is a *Euclidean space*. Following the established tradition, it will be designated l^2.

We introduce in l^2, as in any Euclidean space, the norm of every element $x = (x_1, x_2, \ldots, x_n, \ldots)$, setting it equal to

$$\| x \| = \sqrt{(x, x)} = \sqrt{\sum_{k=1}^{\infty} x_k^2}. \tag{11.2}$$

(Since the series (11.1) is convergent, such a definition is meaningful).

As usual, two elements of l^2 are said to be orthogonal if their scalar product is equal to zero.

Recall that an *orthonormal system* in a Euclidean space is a sequence of elements $\{e_k\}$ of that space satisfying two requirements: (1) any

---

* The convergence of the series $\sum_{k=1}^{\infty} (x_k + y_k)^2$ follows immediately from the inequality $(x_k + y_k)^2 \leqslant 2x_k^2 + 2y_k^2$ and from the convergence of $\sum_{k=1}^{\infty} x_k^2$ and $\sum_{k=1}^{\infty} y_k^2$.

** The formulation of the axioms of a linear space can be found in any course of linear algebra.

*** The convergence of this series follows from the inequality $|x_k y_k| \leqslant \leqslant \frac{1}{2}(x_k^2 + y_k^2)$ and from the convergence of $\sum_{k=1}^{\infty} x_k^2$ and $\sum_{k=1}^{\infty} y_k^2$.

two elements of the sequence should be orthogonal; (2) the norm of each element should be equal to unity.

We prove that there is a *closed* (and therefore, by Theorem 10.7, *complete*) orthonormal system in l^2*. We show that such a system is a sequence of elements

$$\left.\begin{aligned} e_1&=(1,\ 0,\ 0,\ \ldots,\ 0,\ \ldots),\\ e_2&=(0,\ 1,\ 0,\ \ldots,\ 0,\ \ldots),\\ e_3&=(0,\ 0,\ 1,\ \ldots,\ 0,\ \ldots),\\ &\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{11.3}$$

That this system is orthonormal is obvious (the norm (11.2) for every element e_k is equal to unity and the scalar product of any two elements is an infinite sum of products each of which is equal to zero). To prove the closure of the orthonormal system (11.3) it suffices to establish that for any element $x=(x_1,\ x_2,\ \ldots,\ x_n,\ \ldots)$ of l^2 its Fourier series with respect to the system (11.3) converges to it in the norm of l^2**.

Since the Fourier coefficients $(x,\ e_k)$ of an element x coincide with its coordinates x_k, the nth partial sum of its Fourier series is equal to $\sum\limits_{k=1}^{n} x_k e_k$ and it is sufficient for us to prove that

$$\lim_{n\to\infty} \Big\| \sum_{k=1}^{n} x_k e_k - x \Big\| = 0. \tag{11.4}$$

But from the definition of the norm (11.2), from the orthonormality of a system $\{e_k\}$ and from the properties of a scalar product it follows that

$$\begin{aligned} \Big\| \sum_{k=1}^{n} x_k e_k - x \Big\|^2 &= \left(\sum_{k=1}^{n} x_k e_k - x,\ \sum_{k=1}^{n} x_k e_k - x \right) = \\ &= \sum_{k=1}^{n} x_k^2 - 2\sum_{k=1}^{n} x_k (e_k,\ x) + \|x\|^2 = \|x\|^2 - \sum_{k=1}^{n} x_k^2 = \\ &= \sum_{k=1}^{\infty} x_k^2 - \sum_{k=1}^{n} x_k^2 = \sum_{k=n+1}^{\infty} x_k^2, \end{aligned}$$

so that the relation (11.4) follows from the convergence of the series (11.1).

11.1.2. The general form of a linear functional in l^2. We shall discuss functions whose independent variables are elements of l^2

---

* For the definitions of completeness and closure of an orthonormal system see Section 10.2.

** For then any element x of l^2 can be approximated to an unlimited precision in the norm of l^2 by partial sums of that Fourier series.

and whose values are real numbers. Such functions are usually called *functional* (defined in a space l^2).

More precisely, our aim is to make a detailed study of the simplest functional defined in l^2, the so-called *linear* functional.

Definition 1. A functional $l(x)$ defined in a space l^2 is said to be linear if for any elements x and y of l^2 and any real numbers α and β

$$l(\alpha x + \beta y) = \alpha l(x) + \beta l(y).$$

Let x_0 be an arbitrary element of l^2. To geometrize the terminology we shall often call the element x_0 a *point* of l^2.

Definition 2. An arbitrary functional $l(x)$ defined in l^2 is said to be continuous at a point x_0 of l^2 if for any sequence of elements $\{x_n\}$ of l^2 converging in the norm of l^2 to the element x_0 the number sequence $l(x_n)$ converges to $l(x_0)$.

Definition 3. A functional $l(x)$ is said to be continuous *if it is continuous at each point x of l^2.*

Notice at once that *in the case of a linear functional $l(x)$ continuity at least at one point x_0 implies continuity at each point x of l^2*. Indeed, let a linear functional be continuous at a point x_0 and let x be an arbitrary point of l^2. Denote by $\{x_n\}$ an arbitrary sequence of elements of l^2 converging in the norm of l^2 to x. Then the sequence $\{x_0 + x_n - x\}$ converges in the norm of l^2 to x_0 and from the continuity of the functional at x_0 it follows that

$$l(x_0 + x_n - x) \to l(x_0) \quad \text{when} \quad n \to \infty. \tag{11.5}$$

But from the linearity of the functional it follows that $l(x_0 + x_n - x) = l(x_0) + l(x_n) - l(x)$. From the last equation and from (11.5) we get $l(x_n) \to l(x)$ as $n \to \infty$, which just means that the functional is continuous at x.

Definition 4. A functional $l(x)$ is said to be bounded if there is a constant C such that for all elements x of l^2

$$|l(x)| \leqslant C \|x\|. \tag{11.6}$$

Theorem 11.1. For a linear functional $l(x)$ to be continuous it is necessary and sufficient that it should be bounded.

Proof. (1) *Necessity*. Let a linear functional $l(x)$ be continuous. Suppose there is no constant C satisfying inequality (11.6). Then there is a sequence of nonzero elements x_n^* such that $|l(x_n)| \geqslant n^2 \|x_n\|$. Set $y_n = \frac{1}{n\|x_n\|} x_n$. Since $\|y_n - 0\| = \|y_n\| = \frac{1}{n} \to 0$ as $n \to \infty$, by the continuity of the functional $l(y_n) \to l(0) = 0$ as

* For a zero element 0 inequality (11.6) is true for any constant C, for by the linearity of the functional $l(0) = l(0x) = 0 \cdot l(x) = 0$.

$n\to\infty$, but this contradicts the inequality $|\,l(y_n)\,| = \frac{1}{n\,\|x_n\|}\times$ $\times\,|\,l(x_n) \geqslant n$. Thus the necessity is proved.

(2) *Sufficiency*. Let a linear functional $l(x)$ be bounded, i.e. let there be a constant C such that for all elements x inequality (11.6) is true. Also let x_0 be an arbitrary point of l^2 and let $\{x_n\}$ be an arbitrary sequence of elements of l^2 converging in the norm of l^2 to x_0. Then by the linearity of the functional $l(x_n) - l(x_0) = l(x_n - x_0)$, so that on the basis of (11.6) $|\,l(x_n) - l(x_0)\,| = |\,l(x_n - x_0)\,| \leqslant C\,\|x_n - x_0\|$. From the last inequality it follows that $l(x_n) \to l(x_0)$ as $n\to\infty$. Thus the sufficiency is proved.

The theorem proved allows us to introduce the norm of a linear continuous functional.

Definition 5. The **norm** *of a linear continuous functional $l(x)$ is the supremum of a relation $\frac{|\,l(x)\,|}{\|x\|}$ on the set of all elements x of a space l^2.*

It will be designated $\|\,l\,\|$.

So by definition

$$\|l\| = \sup_{x\in l^2}\frac{|\,l(x)|}{\|x\|}. \tag{11.7}$$

The following main theorem is true.

Theorem 11.2 (Riesz's theorem). For every linear continuous functional $l(x)$ there is one and only one element a of l^2 such that for all elements x of l^2

$$l(x) = (a, x), \tag{11.8}$$

with $\|\,l\,\| = \|\,a\,\|$.

Proof. Let $\{e_k\}$ be a closed orthonormal system (11.3), $a_k = l(e_k)$ $(k=1, 2, \ldots)$. We show that the sequence of real numbers $(a_1, a, \ldots, a_n)$ is an element of l^2, i.e. show that the series $\sum\limits_{k=1}^{\infty} a_k^2$ converges.

For any n, set $S_n = \sum\limits_{k=1}^{n} a_k e_k$.

Then by the linearity of the functional

$$l(S_n) = \sum_{k=1}^{n} a_k l(e_k) = \sum_{k=1}^{n} a_k^2 = \|\,S_n\,\|^2. \tag{11.9}$$

On the other hand, from Theorem 11.1 and from the definition of the norm of a linear continuous functional (11.7) it follows that

$$|\,l(S_n)\,| \leqslant \|\,l\,\|\cdot\|S_n\,\|. \tag{11.10}$$

From (11.9) and (11.10) we get $\| S_n \| \leqslant \| l \|$ or equivalently

$$\sum_{h=1}^{n} a_h^2 \leqslant \| l \|^2. \tag{11.11}$$

The last inequality true for any n proves that the series $\sum_{h=1}^{\infty} a_h^2$ converges, i.e. that the sequence $(a_1, a_2, \ldots, a_n, \ldots)$ is some element of l^2 which we denote by a.

Now let $x = (x_1, x_2, \ldots, x_n, \ldots)$ be an arbitrary element of l^2. Then by the closure of an orthonormal system (11.3) a partial sum of a Fourier series $\sum_{h=1}^{n} x_h e_h$ converges in the norm of l^2 to x as $n \to \infty$. By the continuity of the functional it follows that

$$l\Big(\sum_{h=1}^{n} x_h e_h\Big) \to l(x) \text{ as } n \to \infty.$$

But from the linearity of the functional and from the equation $a_h = l(e_h)$ we get

$$l\Big(\sum_{h=1}^{n} x_h e_h\Big) = \sum_{h=1}^{n} x_h l(e_h) = \sum_{h=1}^{n} x_h a_h.$$

So we have proved that

$$\lim_{n \to \infty} \sum_{h=1}^{n} x_h a_h = l(x),$$

but this means that we have established equation (11.8) with a *uniquely* determined element a whose coordinates are equal to $l(e_h)$.

It remains to show that $\| l \| = \| a \|$. It is immediate from inequality (11.11) true for any n that

$$\| a \| \leqslant \| l \|. \tag{11.12}$$

On the other hand, from equation (11.8) we have already proved and from the Cauchy-Buniakowski inequality* $|(a, x)| \leqslant \| a \| \cdot \| x \|$ we get $|l(x)| \leqslant \| a \| \cdot \| x \|$ from which by the definition of the norm (11.7) it follows that

$$\| l \| \leqslant \| a \|. \tag{11.13}$$

From (11.12) and (11.13) we conclude that $\| l \| = \| a \|$. Thus the proof of the theorem is complete.

The theorem establishes a general form of any linear continuous functional in l^2.

---

* By Theorem 10.1 the Cauchy-Buniakowski inequality is true for any two elements of any Euclidean space.

11.1.3. On the weak compactness of a set bounded in the norm of l^2.

Definition 1. A set E of elements of l^2 is said to be bounded (*or* bounded in the norm) *if there is a constant M of elements such that $\|x\| \leqslant M$ for all elements x of E.*

Definition 2. An infinite set E of elements of l^2 is said to be compact *if we can choose a subsequence $\{x_{n_k}\}$ of any sequence of elements $\{x_n\}$ of E converging in the norm of l^2.*

It is obvious that *any compact set E of l^2 elements is bounded**.

In a *finite*-dimensional Euclidean space the converse is also true: *any bounded set E containing an infinite number of elements is compact* (the Bolzano-Weirstrass theorem). But in an *infinite*-dimensional space, such as l^3, the boundedness of an infinite set of elements of E no longer implies its compactness.

For instance, the set $\{e_k\}$ of all the elements of an orthonormal system (11.3) is bounded (for the norms of all elements are equal to unity), but it is not compact (since for a sequence of elements to converge in the norm of l^2 it is necessary that the norm of the difference of two elements with integers k and $k+1$ should converge to zero as $k \to \infty$, and for any subsequence of the elements of (11.3) $\| e_k - e_l \|^2 = \| e_k \|^2 + \| e_l \|^2 = 2$ given any k and l not equal to each other).

It is natural to try to introduce the concept of compactness of a set in a weaker sense (than in Definition 2) so that any bounded set (containing an infinite number of elements) should turn out to be compact in that weak sense.

Definition 3. A sequence $\{x_n\}$ of elements of a space l^2 is said to converge weakly *to an element x_0 of that space if for any element a of l^2*

$$(x_n, a) \to (x_0, a) \quad \text{as} \quad n \to \infty.$$

Notice that the convergence of $\{x_n\}$ to x_0 in the norm of l^2 and the Cauchy-Buniakowski inequality imply the weak convergence of $\{x_n\}$ to x_0, since $|(x_n, a) - (x_0, a)| = |(x_n - x_0, a)| \leqslant \leqslant \sqrt{\| x_n - x_0 \| \cdot \| a \|}$ for any element a. The weak convergence of $\{x_n\}$ to x_0 does not in general imply the convergence of $\{x_n\}$ to x_0 in the norm of l^2. For instance, the sequence $\{e_k\}$ of all the elements of an orthonormal system (11.3) converges weakly to a zero element 0, because for any element a of l^2 we have the Bessel inequality** $\sum\limits_{k=1}^{\infty} (e_k, a)^2 \leqslant \| a \|^2$ according to which $(e_n, a) \to (0, a) = 0$ as

* Indeed, the boundedness of E would imply the existence of a sequence of elements of E for which the sequence of norms is infinitely large. Any subsequence of such a sequence diverges in the norm of l^3, which contradicts the compactness condition of E.

** By Theorem 10.4 the Bessel inequality is true for every element and any orthonormal system in an arbitrary Euclidean space.

$n \to \infty$. Moreover, it has been proved above that $\{e_k\}$ does not converge in the norm of l^2.

Convergence in the norm of l^2 (in contrast to weak convergence) is often called *strong convergence.*

Definition 4. An infinite set E of elements of l^2 is said to be weakly compact *if we can choose a weakly convergent subsequence of any sequence of elements $\{x_n\}$ of E.*

The following fundamental theorem is true.

Theorem 11.3. Any bounded set of an infinite number of elements in l^2 is weakly compact.

Proof. Let E be an arbitrary bounded subset of l^2 containing an infinite number of elements and let $\{x_n\}$ be an arbitrary sequence of elements of E. The condition of boundedness of E allows us to state that $\| x_n \| \leqslant M$, where M is some constant. But then it follows from $\| x_n \|^2 = \sum\limits_{k=1}^{\infty} x_{nk}^2$ that for any k the number sequence of the kth coordinates x_{nk} of the elements x_n is bounded. Therefore, by the Bolzano-Weierstrass theorem (see Theorem 3.3 in [1]) we can choose from $\{x_n\}$ a subsequence of elements $\{x_n^{(1)}\}$ such that the first coordinates of those elements form a convergent number sequence and then we can choose from $\{x_n^{(1)}\}$ a subsequence of elements $\{x_n^{(2)}\}$ such that both the first and the second coordinates of those elements form convergent number sequences and so on. After k steps we choose a subsequence of elements $\{x_n^{(k)}\}$ in which each of the first k coordinates form a convergent number sequence.

Set $y_n = x_n^{(n)}$. It is obvious that $\{y_n\}$ is a subsequence of the original sequence of elements $\{x_n\}$ and that the sequence formed by *any* coordinate of the elements y_n is a convergent number sequence, i.e. if $y_n = (y_{n1}, y_{n2}, \ldots, y_{nk}, \ldots)$ then for every k the sequence of y_{nk} converges as $n \to \infty$. We denote by ξ_k the limit of a sequence of the kth coordinates of elements y_n, i.e. set $\xi_k = \lim\limits_{n \to \infty} y_{nk}$ $(k = 1, 2, \ldots)$ and show that the sequence $(\xi_1, \xi_2, \ldots, \xi_k, \ldots)$ is some element of l^2, i.e. show that the series $\sum\limits_{k=1}^{\infty} \xi_k^2$ converges. Since $\| y_n \| \leqslant \leqslant M$ for every n, we have for every n

$$\sum_{k=1}^{\infty} y_{nk}^2 \leqslant M^2 \tag{11.14}$$

and clearly

$$\sum_{k=1}^{N} y_{nk}^2 \leqslant M^2 \tag{11.15}$$

(for *any* fixed N and for every n).

Proceeding in (11.15) to the limit as $n \to \infty$ we get $\sum_{k=1}^{N} \xi_k^2 \leqslant M^2$ for any N, and this means that the sequence $(\xi_1, \xi_2, \ldots, \xi_k, \ldots)$ is some element of l^2 which we denote by ξ.

It remains to prove that $\{y_n\}$ weakly converges to that element ξ, i.e. that for any element $a = (a_1, a_2, \ldots, a_k, \ldots)$ of l^2 $\lim_{n\to\infty} (y_n, a) = (\xi, a)$ or equivalently

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} y_{nk} a_k = \sum_{k=1}^{\infty} \xi_k a_k.$$

By virtue of the fact that $\lim_{n\to\infty} y_{nk} = \xi_k$ and by the theorem on passage to the limit term by term (see Theorem 1.6) it suffices to prove that the series

$$\sum_{k=1}^{\infty} y_{nk} \cdot a_k \tag{11.16}$$

converges uniformly with respect to every n. Fix an arbitrary $\varepsilon > 0$. The convergence of $\sum_{k=1}^{\infty} a_k^2$ implies the existence of m_0 such that

$$\sum_{k=m+1}^{m+p} a_k^2 < \frac{\varepsilon^2}{M^2} \tag{11.17}$$

for every $m \geqslant m_0$ and every natural p $(p = 1, 2, \ldots)$.

Applying to the remainder of the series (11.16) the Cauchy-Buniakowski inequality for the sums*

$$\left| \sum_{k=m+1}^{m+p} y_{nk} a_k \right| \leqslant \left[\sum_{k=m+1}^{m+p} y_{nk}^2 \sum_{k=m+1}^{m+p} a_k^2 \right]^{1/2}$$

and using inequalities (11.14) and (11.17) we get

$$\left| \sum_{k=m+1}^{m+p} y_{nk} a_k \right| < \varepsilon$$

for every $m \geqslant m_0$, every natural p and for every n at once. But this means that (11.16) converges uniformly with respect to every n. Thus the theorem is proved.

It has numerous applications. In particular, it is widely used in the theory of variational methods of solving problems in mathematical physics.

---

* This inequality has been established in Supplement 1 to Chapter 10 of [1].

11.2. THE SPACE L^2

11.2.1. The simplest properties of the space L^2. We are already familiar with the space L^2 from the study of classes L^p, with $p \geqslant 1$, in Section 8.4.7.

Recall that a space $L^2(E)$ is the set of all functions $\{f(x)\}$ such that every function $f(x)$ is measurable on a set E and every function $f^2(x)$ is summable (i.e. integrable in the Lebesgue sense) on E. We do not distinguish between functions equivalent on E, considering them to be a single element of $L^2(E)$.

$L^2(E)$ is briefly called a *space of functions with a square summable* (on a set E).

Note at once that all the integrals in this section are understood in the Lebesgue sense and that a set E is understood to be a measurable set of positive finite measure on an infinite line, although the entire theory to be presented could be extended without any complications to include the case of an arbitrary set of positive measure E in a space of any number n of dimensions.

It was established in Section 8.4.7 that a space $L^2(E)$ is a normalized linear space with the norm of any element $f(x)$ of the form

$$\|f\| = \left(\int_E f^2(x)\,dx\right)^{1/2}. \tag{11.18}$$

The space $L^2(E)$ differs substantially from all other spaces $L^p(E)$, with $p \neq 2$, in that $L^2(E)$ is a Euclidean space with a scalar product of any two elements $f(x)$ and $g(x)$ of the form*

$$(f, g) = \int_E f(x)\,g(x)\,dx. \tag{11.19}$$

The validity in $L^2(E)$ of all the four axioms of a scalar product** easily follows from the independence of the product $f(x)\,g(x)$ of the order of the factors, from the linear properties of the integral and from the equivalence to zero of a measurable, summable and nonnegative function $f^2(x)$.

Also notice that from (11.18) and (11.19) it follows that in L^2 (as in any Euclidean space) the norm and scalar product are related by

$$\|f\| = \sqrt{(f, f)}.$$

---

* For the definition of a Euclidean space and a scalar product see Section 10.1.

** The axioms of a scalar product can be found in Section 10.1.

Finally, recall that it has been proved in Section 8.4.7 that the space $L^2(E)$ is complete*.

We now proceed to show the more profound properties of $L^2(E)$.

11.2.2. The separability of the space L^2. First consider an arbitrary normalized linear space R.

Definition 1. A set M of elements of a normalized linear space R is said to be everywhere dense (*or* dense *in R*) *if for any element f of R we can choose a sequence of elements $\{f_n\}$ of M converging in the norm of R to f.*

Definition 2. A normalized linear space R is said to be separable if in it there is a dense set of elements M countable everywhere.

Our aim here is to prove the separability of L^2.

Theorem 11.4. A set of functions continuous on E is everywhere dense in $L^2(E)$.

Proof. Let $f(x)$ be an arbitrary function in $L^2(E)$. We may assume without loss of generality that $f(x) \geqslant 0$. Introducing two nonnegative functions

$$f^+(x)=\frac{1}{2}(|f(x)|+f(x)),\quad f^-(x)=\frac{1}{2}(|f(x)|-f(x)),$$

it is indeed easy to see that the theorem is true for any function $f \in L^2$ provided it has been proved for nonnegative functions.

In addition we may assume that $f(x)$ takes on finite values everywhere. So let $f(x) \in L^2(E)$ and $0 \leqslant f(x) < \infty$.

Consider for every n a sequence of disjoint sets**

$$E_n^k=E\left[\frac{k}{2^n}\leqslant f(x)<\frac{k+1}{2^n}\right]\quad (k=0,\ 1,\ 2,\ \ldots).$$

Then obviously for any n $(n = 1, 2, \ldots)$ the sum of those sets over all $k = 0, 1, \ldots$ yields a set E, i.e. $E = \bigcup\limits_{k=0}^{\infty} E_n^k$.

Construct a sequence $\{f_n(x)\}$ of functions defined on E for every n, by setting $f_n(x) = k/2^n$ for x that is in E_n^k. Thus every function $f_n(x)$ is a "step"-function on E (taking on at most a countable number of values).

It is also obvious that for every n and every point x of E

$$0\leqslant f(x)-f_n(x)<1/2^n,$$

whence it follows that $\{f_n(x)\}$ converges to $f(x)$ uniformly on E. Set $\Psi_n(x) = \min\{n, f_n(x)\}$.

---

* Recall that a normalized linear space R is said to be *complete* if for any fundamental sequence $\{f_n\}$ of elements of that space (i.e. for a sequence $\{f_n\}$ for which $\lim\limits_{n\to\infty} \sup\limits_{m\geqslant n} \|f_m - f_n\| = 0$) there is an element f of R to which that sequence converges in R.

** Recall that the symbol E [f satisfies condition A] designates a set of all points of E for which $f(x)$ satisfies condition A.

Every function $\Psi_n(x)$ takes on on E only a *finite* number of values, the sequence $\{\Psi_n(x)\}$ converging to $f(x)$ *everywhere* on E. Since, in addition, $0 \leqslant f(x) - \Psi_n(x) \leqslant f(x)$ everywhere on E, whence it follows that $[f(x) - \Psi_n(x)]^2 \leqslant f^2(x)$, by the corollary of Theorem 8.19 the sequence $[f(x) - \Psi_n(x)]^2$ converges to zero in $L^1(E)$, i.e. the sequence of $\Psi_n(x)$ converges to $f(x)$ in $L^2(E)$.

It remains to prove that every function $\Psi_n(x)$ can be approximated in the norm of $L^2(E)$ by a continuous function to an unlimited precision. Recall that every function $\Psi_n(x)$ takes on only a finite number of values, i.e. has the form $\Psi_n(x) = \sum_{k=1}^{m} a_k \omega_k(x)$, where a_k $(k = 1, 2, \ldots, m)$ are constant numbers and $\omega_k(x)$ are the so-called *characteristic* functions of the sets E_k:

$$\omega_k(x) = \begin{cases} 1 & \text{on a set } E_k, \\ 0 & \text{outside } E_k. \end{cases}$$

Thus, to complete the proof it suffices to construct a sequence of continuous functions converging in $L^2(E)$ to a function $\omega(x)$ of the form

$$\omega(x) = \begin{cases} 1 & \text{on } E_0, \\ 0 & \text{outside } E_0. \end{cases}$$

where E_0 is some measurable set in E.

For a set E_0 and for any n there are an open set G_n containing E_0 and a closed set F_n contained in E_0 such that the measure of the difference $G_n - F_n$ is less than $1/n$*.

Denote by $\widetilde{F}_n$ the complement of G_n and set

$$\varphi_n(x) = \frac{\rho(x, \widetilde{F}_n)}{\rho(x, \widetilde{F}_n) + \rho(x, F_n)},$$

where $\rho(x, F)$ stands for the distance from a point x to a set F.

It is obvious that every function $\varphi_n(x)$ is continuous on E, equal to unity on F_n, equal to zero on $\widetilde{F}_n$, and satisfies everywhere the condition $0 \leqslant \varphi_n(x) \leqslant 1$. Hence we obtain for the norm of the difference $\varphi_n(x) - \omega(x)$ the following estimate:

$$\|\varphi_n - \omega\|^2_{L^2(E)} = \int_E [\varphi_n(x) - \omega(x)]^2\, dx \leqslant \int_{G_n - F_n} dx < \frac{1}{n} \qquad (11.20)$$

which completes the proof.

We now prove the following *main* theorem.

Theorem 11.5. For any bounded measurable set E a space $L^2(E)$ is separable.

* By the definition of the measurability of a set E_0 and the corollary of Theorem 8.5 (see Section 8.2.2).

Proof. We first prove the case where E is a closed interal $[a, b]$. We prove that in this case we may take the set M of all polynomials with rational coefficients as a countable everywhere dense set in $L^2 [a, b]$*.

According to Theorem 11.4 any function $f(x)$ in $L^2([a, b])$ can be approximated in the norm of $L^2([a, b])$ to an unlimited precision by a continuous function. Further, according to Theorem 1.18 (Weierstrass) *any function continuous on $[a, b]$ can be approximated uniformly* on that interval (and therefore in the norm of $L^2([a, b])$) to an unlimited precision by an algebraic polynomial with real coefficients.

Finally, it is obvious that an algebraic polynomial with real coefficients can be approximated uniformly on $[a, b]$ and therefore in the norm of $L^2([a, b])$ to an unlimited precision by a polynomial with rational coefficients. This completes the proof for the case where E is a closed interval $[a, b]$.

Now let E be an arbitrary bounded measurable set. Since E is bounded, there is a closed interval $[a, b]$ containing E.

Let $f(x)$ be an arbitrary function in $L^2(E)$. Extend it to $[a, b]$ by setting it equal to zero outside E. It remains to notice that the function $f(x)$ thus extended is in $L^2([a, b])$ and therefore, by what has been proved above, can be approximated to an unlimited precision in the norm of $L^2([a, b])$ (and the more so in the norm of $L^2(E)$) by polynomials with rational coefficients. In this case too, therefore, polynomials with rational coefficients form a set dense everywhere in $L^2(E)$. This completes the proof.

11.2.3. The existence in L^2 of a closed orthonormal system containing a countable number of elements. To construct in L^2 a closed orthonormal system of elements we shall assume the existence in L^2 of a countable, everywhere dense set of elements $f_1, f_2, \ldots, f_n, \ldots$.

We shall prove that a closed orthonormal system can be constructed using finite linear combinations** of elements of the everywhere dense set $f_1, f_2, \ldots, f_n, \ldots$.

Such a method of constructing an orthonormal system is usually called *orthogonalization process*.

We shall assume that there are no linearly dependent*** elements among $f_1, f_2, \ldots, f_n, \ldots$ (otherwise by successively increasing n we should remove out of $\{f_n\}$ every element f_n which is a linear combination of $f_1, f_2, \ldots, f_{n-1}$).

---

* That such a set M is countable follows from the countability of all rational numbers and from the countability of the number of all polynomials of different degree.

** An element Ψ_n is said to be a linear combination of elements $f_1, f_2, \ldots, f_m$ if there are real numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that $\Psi_n = \alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_m f_m$.

*** This means that none of the elements f_n of $\{f_k\}$ is a linear combination of a finite number of other elements of $\{f_k\}$.

We construct a system of mutually orthogonal nonzero elements $\Psi_1, \Psi_2, \ldots, \Psi_n, \ldots$ such that for any n each of the elements $\Psi_1, \Psi_2, \ldots, \Psi_n$ is a linear combination of $f_1, f_2, \ldots, f_n$ and, conversely, each of the elements $f_1, f_2, \ldots, f_n$ is a linear combination of $\Psi_1, \Psi_2, \ldots, \Psi_n$*.

We prove by mathematical induction that the system of elements $\Psi_1, \Psi_2, \ldots, \Psi_n, \ldots$ can be successively defined using the relations

$$\Psi_1 = f_1 \tag{11.21}$$

$$\Psi_n = \begin{vmatrix} (f_1, \Psi_1) & (f_1, \Psi_2) & \ldots & (f_1, \Psi_{n-1}) & f_1 \\ (f_2, \Psi_1) & (f_2, \Psi_2) & \ldots & (f_2, \Psi_{n-1}) & f_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (f_n, \Psi_1) & (f_n, \Psi_2) & \ldots & (f_n, \Psi_{n-1}) & f_n \end{vmatrix} \quad \text{for} \quad n \geqslant 2. \tag{11.22}$$

Clearly, the element Ψ_1 defined by (11.21) is nonzero (for otherwise for any n the elements $f_1, f_2, \ldots, f_n$ would turn out to be linearly dependent).

Thus for $n = 1$ all the above requirements are fulfilled. We now suppose that the system $\Psi_1, \Psi_2, \ldots, \Psi_{n-1}$ constructed using relations (11.21) and (11.22) satisfies all the above requirements and shows that then the system $\Psi_1, \Psi_2, \ldots, \Psi_n$ constructed using the same relations satisfies them as well.

From (11.22) it is clear that an element Ψ_n is some linear combination of $f_1, f_2, f_3, \ldots, f_n$ and is thus nonzero (otherwise that linear combination would turn out to be a zero element, i.e. the elements $f_1, f_2, \ldots, f_n$ would turn out to be linearly dependent).

Further, since the elements $f_1, f_2, \ldots, f_{n-1}$ can be linearly expressed in terms of $\Psi_1, \Psi_2, \ldots, \Psi_{n-1}$ and since the minor in the right-hand bottom corner of the determinant (11.22) of f_n is equal to $\|\Psi_{n-1}\|$** and hence nonzero, it follows from (11.22) that f_n can be linearly expressed in terms of $\Psi_1, \Psi_2, \ldots, \Psi_n$.

Finally, from (11.22) it follows immediately that the element Ψ_n is orthogonal to each of the elements $\Psi_1, \Psi_2, \ldots, \Psi_{n-1}$. Indeed, if k is any of the integers $1, 2, \ldots, n-1$, then performing scalar multiplication of both sides of (11.22) by Ψ_k we obtain on the right side a determinant whose kth and nth columns are the same. It follows from the equality of such a determinant to zero that $(\Psi_n, \Psi_k) = 0$ for all $k = 1, 2, \ldots, n-1$.

This completes the induction and thus the system $\Psi_1, \Psi_2, \ldots$ $\ldots, \Psi_n, \ldots$ satisfying the above requirements is constructed.

---

* In terms of linear algebra this means that the span of $\Psi_1, \Psi_2, \ldots, \Psi_n$ coincides with the span of $f_1, f_2, \ldots, f_n$.

** To show this it suffices to write equation (11.22) for $(n-1)$ and perform the scalar multiplication of it by Ψ_{n-1}.

On setting now for every n $\varphi_n = \Psi_n / \| \Psi_n \|$ we obtain an orthonormal system $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$.

The closure of the constructed system $\{\varphi_n\}$ follows immediately from the fact that each element of the everywhere dense set $\{f_n\}$ is a linear combination of a finite number of elements of $\{\varphi_n\}$.

From the countability of the everywhere dense set of elements $f_1, f_2, \ldots, f_n, \ldots$ it follows that the closed orthonormal system constructed contains *at most a countable number* of elements. But the number of elements of that system *cannot be finite*, for this would mean that the space L^2 is finite-dimensional*.

This completely proves the existence in L^2 of a closed orthonormal system consisting of a countable number of elements.

We note in conclusion that a closed orthonormal system of elements of L^2 is often called an *orthonormal basis***.

11.2.4. Isomorphism of the spaces L^2 and l^2 and its consequences. In $L^2(E)$, just as in l^2, the concepts of weak convergence of a sequence of elements and of weak compactness of a set of elements can be introduced.

Definition 1. A sequence $\{f_n(x)\}$ of elements of a space $L^2(E)$ is said to converge weakly *to an element $f(x)$ of that space if for any element $g(x)$ of $L^2(E)$*

$$(f_n, g) \to (f, g) \quad \text{as} \quad n \to \infty$$

or equivalently

$$\int_E f_n(x)\, g(x)\, dx \to \int_E f(x)\, g(x)\, dx \quad \text{as} \quad n \to \infty.$$

It can be proved, as elementary as for the case of l^2, that convergence of $\{f_n(x)\}$ to $f(x)$ in the norm of $L^2(E)$ implies weak convergence of $\{f_n(x)\}$ to $f(x)$. Of course, the weak convergence of the elements of $L^2(E)$ does not imply their convergence in the norm of $L^2(E)$ (any orthonormal sequence of elements of $L^2(E)$ may serve as an example).

Definition 2. An infinite set M of elements of $L^2(E)$ is said to be weakly compact *if we can choose a weakly convergent subsequence of any sequence of elements $\{f_n(x)\}$ in M.*

The concept of linear continuous functional can be introduced in L^2 in a way quite similar to that for l^2.

---

* That the dimensionality of $L^2(E)$ is equal to infinity follows immediately from the fact that for any preassigned n there are n linearly independent elements $1, x, x^2, \ldots, x^{n-1}$ in $L^2(E)$.

** A system of elements $\{\varphi_n\}$ is said to be a *basis* of $L^2(E)$ if for any element f of $L^2(E)$ there is a unique expansion of that element into a series $\sum_{n=1}^{\infty} c_n \varphi_n$ with constant coefficients c_n converging to f in the norm of $L^2(E)$.

Definition 3. A functional $l(f)$ defined on the elements of f of $L^2(E)$ is said to be **linear** *if for any two elements f and g of $L^2(E)$ and for any real numbers α and β, $l(\alpha f+\beta g)=\alpha l(f)+\beta l(g)$.* Let us agree to call elements f, whenever convenient, *points* of $L^2(E)$.

Definition 4. A functional $l(f)$ defined on the elements of $L^2(E)$ is said to be **continuous** at a point f_0 *of $L^2(E)$ if for any sequence $\{f_n\}$ of elements of $L^2(E)$ converging in the norm of $L^2(E)$ to the element f_0 the number sequence $l(f_n)$ converges to $l(f_0)$.*

Definition 5. A functional $l(f)$ is said to be **continuous** *if it is continuous at each point f of $L^2(E)$.*

It is easy to prove, as for the case of l^2, that if a linear functional of $L^2(E)$ is continuous at least at one point of $L^2(E)$, then it is continuous everywhere in $L^2(E)$, i.e. simply continuous.

The question naturally arises as to where it is possible to extend Theorem 11.2 on the general form of a linear continuous functional and Theorem 11.3 on weak compactness of any set bounded (in the norm) to include the case of $L^2(E)$.

We shall establish a close connection between L^2 and l^2 which will allow us to immediately prove that both theorems are valid for L^2.

We introduce the following fundamental notion.

Definition 6. Two arbitrary Euclidean spaces R and R' are said to be **isomorphic** *if a one-to-one correspondence can be established between the elements of R and R' so that, provided the elements x' and y' of R' are the images of elements x and y of R, the following requirements are fulfilled:* (1) *the element $x'+y'$ of R' is the image of the element $x+y$ of R;* (2) *for any real λ the element $\lambda x'$ of R' is the image of the element λx of R;* (3) *the scalar products (x', y') and (x, y) are equal to each other.*

It is established in linear algebra that all n-dimensional Euclidean spaces are isomorphic to one another and to the space E^n.

The principal aim of Section 11.2.4 is to establish that infinite-dimensional Euclidean spaces $L^2(E)$ and l^2 are isomorphic. But we shall first prove the following remarkable theorem.

*Theorem 11.6 (the Riesz-Fisher theorem). Let $\{\varphi_n\}$ be an arbitrary orthonormal system in $L^2(E)$**. *Then for any sequence of real numbers $(c_1, c_2, \ldots, c_n, \ldots)$ that satisfies the condition $\sum\limits_{k=1}^{\infty} c_k^2<\infty$, i.e. is an element of l^2, there is a unique function $f(x)$ in $L^2(E)$ such that $c_n=(f, \varphi_n)=\int\limits_E f(x)\varphi_n(x)\,dx$ and $\sum\limits_{k=1}^{\infty} c_k^2=$ $=\|f\|^2=\int\limits_E f^2(x)\,dx$.*

* Neither the completeness nor, clearly, the closure of the system is assumed.

Proof. Set $f_n = \sum_{k=1}^{n} c_k \varphi_k$. The sequence $\{f_n\}$ is fundamental, since for $m \geqslant n$ $\| f_m - f_n \|^2 = \sum_{k=n+1}^{m} c_k^2$ and, as stated, the series $\sum_{k=1}^{\infty} c_k^2$ converges. But then by the completeness of $L^2(E)$ (established in Section 8.4.7) there is an element f of $L^2(E)$ such that

$$\lim_{n\to\infty} \| f_n - f \| = \lim_{n\to\infty} \Big\| \sum_{k=1}^{n} c_k \varphi_k - f \Big\| = 0. \tag{11.23}$$

From the last relation and from the Bessel identity (10.17) established in Section 10.1* it follows that

$$\lim_{n\to\infty} \sum_{k=1}^{n} c_k^2 = \| f \|^2, \quad \text{i.e.} \quad \sum_{k=1}^{\infty} c_k^2 = \| f \|^2.$$

We prove that $(f, \varphi_k) = c_k$ for any k. To do this notice that by *the orthonormality of* $\{\varphi_k\}$, *for all* $n \geqslant k$

$$(f_n, \varphi_k) = \Big(\sum_{l=1}^{n} c_l \varphi_l, \ \varphi_k \Big) = \sum_{l=1}^{n} c_l (\varphi_l, \varphi_k) = c_k, \tag{11.24}$$

and that by the Cauchy-Buniakowski inequality

$$| (f_n, \varphi_k) - (f, \varphi_k) | = | (f_n - f, \varphi_k) | \leqslant$$
$$\leqslant \sqrt{\| f_n - f \| \cdot \| \varphi_k \|} = \sqrt{\| f_n - f \|}$$

and by (11.23)

$$(f_n, \varphi_k) \to (f, \varphi_k) \text{ as } n \to \infty. \tag{11.25}$$

From (11.24) and (11.25) we get $(f, \varphi_k) = c_k$ for any k.

It remains to prove that f is a *single* element of $L^2(E)$ satisfying all the conditions of the theorem. Let g be any other element of $L^2(E)$ satisfying all the conditions of the theorem. From the Cauchy-Buniakowski inequality $| (f_n - f, g) | \leqslant \sqrt{\| f_n - f \|} \times \sqrt{\| g \|}$ and from (11.23) it follows that

$$(f_n - f, g) \to 0 \text{ as } n \to \infty. \tag{11.26}$$

But from the equation $(g, \varphi_k) = c_k$ and from the scalar product axioms it follows that

$$(f_n - f, g) = \Big(\sum_{k=1}^{n} c_k \varphi_k - f, g \Big) = \sum_{k=1}^{n} c_k (g, \varphi_k) -$$
$$-(f, g) = \sum_{k=1}^{n} c_k^2 - (f, g),$$

* This Bessel identity is true for any orthonormal system in an arbitrary Euclidean space.

so by (11.26)

$$\sum_{k=1}^{\infty} c_k^2 = (f, g). \tag{11.27}$$

From (11.27) and from the relations $\sum_{k=1}^{\infty} c_k^2 = \|f\|^2$ and $\sum_{k=1}^{\infty} c_k^2 = \|g\|^2$ we get

$$\|f-g\| = (f-g, f-g) = \|f\|^2 - 2(f, g) + \|g\|^2 = 0.$$

But this means that the difference $f-g$ is a zero element of $L^2(E)$, i.e. $f = g$. This completes the proof.

Remark. If an orthonormal system $\{\varphi\}_n$ is closed or at least complete, then the element f is unique even without the requirement that $\sum_{k=1}^{\infty} c_k^2 = \|f\|^2$ (see in this connection Theorem 10.8).

On the basis of the Riesz-Fisher theorem we prove the following *main* theorem.

Theorem 11.7. $L^2(E)$ and l^2 are isomorphic.

Proof. Choose in $L^2(E)$ a closed orthonormal system $\{\varphi_k\}$ and associate with each element f of $L^2(E)$ an element $c = (c_1, c_2, \ldots, c_n, \ldots)$ of l^2 whose coordinates c_k are of the form $c_k = (f, \varphi_k)$ $(k = 1, 2, \ldots)$. By Theorem 11.6 this is a one-to-one correspondence.

It remains to prove that if corresponding to the elements f and g of $L^2(E)$ are respectively the elements $c = (c_1, c_2, \ldots, c_n, \ldots)$ and $d = (d_1, d_2, \ldots, d_n, \ldots)$ of l^2, then (1) corresponding to the element $f + g$ is the element $c + d = (c_1 + d_1, c_2 + d_2, \ldots, c_n + d_n, \ldots)$, (2) for any λ corresponding to λf is $\lambda c = (\lambda c_1, \lambda c_2, \ldots, \lambda c_n, \ldots)$ (3)

$$(f, g) = (c, d) = \sum_{k=1}^{\infty} c_k d_k \tag{11.28}$$

(this equation is usually called the *generalized Parseval formula*).

(1) and (2) follow immediately from the properties of a scalar product*. We prove (11.28). By the closure of $\{\varphi_k\}$ the following Parseval's formulas are true for f, g, and $f + g$:

$$(f, f) = \sum_{k=1}^{\infty} c_k^2, \quad (g, g) = \sum_{k=1}^{\infty} d_k^2, \tag{11.29}$$

$$(f+g, f+g) = \sum_{k=1}^{\infty} (c_k + d_k)^2. \tag{11.30}$$

* Thus, to prove (1) it suffices to notice that $(f+g, \varphi_k) = (f, \varphi_k) + (g, \varphi_k) = c_k + d_k$.

Subtracting (11.29) from (11.30) we get

$$2(f,\ g)=2\sum_{k=1}^{\infty} c_k d_k.$$

The proof of the theorem is complete.

Theorem 11.7 allows l^2 to be considered as coordinate notation for the elements of $L^2(E)$. It extends all the statements established for l^2 to $L^2(E)$, and vice versa.

In particular, it implies the following.

1°. l^2 is complete.

2°. Any set bounded in the norm of $L^2(E)$ and containing an infinite number of elements of $L^2(E)$ is weakly compact.

3°. For every linear continuous functional $l(f)$ defined on the elements f of $L^2(E)$ there is one and only one element g of $L^2(E)$ such that, for every element f of $L^2(E)$, $l(f)=(f,\ g)$ with

$$\| l \| = \sup_{f\in L^2(E)} \frac{|l(f)|}{\| f \|} = \| g \|.$$

From a quantum-mechanical point of view Theorem 11.7 is a mathematical justification of the equivalence of Heisenberg's "matrix mechanics" and Schrödinger's "wave mechanics", the mathematical formalism of the former being the coordinate space l^2 and of the second—the space L^2 of functions with an integrable square.

Theorem 11.7 naturally suggests that l^2 and L^2 are both but different specific realizations of the same abstract space which we now proceed to consider.

11.3. ABSTRACT HILBERT SPACE

11.3.1. Abstract Hilbert space. A Hilbert space H with two particular forms of which, l^2 and L^2, we have already got acquainted can be introduced axiomatically as a collection of elements $X, Y, Z, \ldots$ of any nature that satisfy a certain system of axioms.

We enumerate all the axioms which must be satisfied by elements of an abstract Hilbert space H.

I. (*a*) Axiom on the existence of a rule which associates with any two elements X and Y of a space H an element Z of that space called the sum of X and Y.

(b) Axiom on the existence of a rule which associates with any element X of H and any real number λ an element of H called a product of X by λ.

(c) Eight axioms of a linear space*.

II. (a) Axiom on the existence of a rule associating with any two elements X and Y of H a number called a scalar product of X and Y and designated (X, Y).

(b) Four axioms of a scalar product**.

III. Axiom on the completeness of a space H with respect to the norm defined by the equation $\| X \| = \sqrt{(X, X)}$***.

IV. Axiom on the existence in H of any preassigned number of linearly independent elements.

V. Axiom on the existence in H of a countable, everywhere dense set (in the sense of the norm of H) of elements.

In other words, a Hilbert space H *is any linear Euclidean complete infinite-dimensional separable space.*

The following notions are introduced in the Hilbert space H: (1) convergence of a sequence of elements *in the norm and weak convergence* (a sequence of elements $\{X_n\}$ is said to converge weakly to an element X if for any Y, $(X_n, Y) \to (X, Y)$ as $n \to \infty$); (2) *weak compactness* of a set M of elements of H (defined as the possibility of choosing a weakly convergent subsequence of any sequence of elements of M); (3) a *linear* and *continuous* functionals $l(X)$ defined on the elements X of H ($l(X)$ is said to be *linear* if $l(\alpha X + \beta Y) = \alpha l(X) + \beta l(Y)$ for any elements X and Y of H and any real numbers α and β; $l(X)$ is said to be *continuous* at a "*point*" X_0 if $l(X)_n \to l(X_0)$ for any sequence $\{X_n\}$ of elements of H for which $\| X_n - X_0 \| \to 0$; $l(X)$ is said to be *continuous* if it is continuous at each point X of H).

The existence of a *closed orthonormal system of elements* $\{\Phi_n\}$ can be proved for the abstract Hilbert space H in a way quite similar

---

* These axioms can be found in any course of linear algebra. They are listed here for convenience.

1°. $X + Y = Y + X$ (for any elements X and Y).

2°. $X + (Y + Z) = (X + Y) + Z$ (for any elements X, Y and Z).

3°. There is an element 0 such that $X + 0 = X$ for any element X.

4°. For every element X there is an element X' such that $X + X' = 0$.

5°. $\alpha(\beta X) = (\alpha\beta)\cdot X$ for any element X and any real numbers α and β.

6°. $1 \cdot X = X$ for any element X.

7°. $(\alpha + \beta) X = \alpha X + \beta X$ for any element X and any real numbers α and β.

8°. $\alpha(X + Y) = \alpha X + \alpha Y$ for any elements X and Y and any real number α.

** Axioms of a scalar product can be found in Section 10.1. They are listed here for convenience.

1°. $(X, Y) = (Y, X)$ for any elements X and Y.

2°. $(X + Y, Z) = (X, Z) + (Y, Z)$ for any elements X, Y, and Z.

3°. $(\alpha X, Y) = \alpha (X, Y)$ for any elements X and Y and for any real number α.

4°. $(X, X) > 0$ for every nonzero element X, $(0, 0) = 0$.

*** For the definition of the completeness of a normalized linear space see Section 8.4.7.

to that used in Section 11.2.3 for L^2 (to do this the orthogonalization process for a countable everywhere dense set of elements of H is carried out).

For the abstract Hilbert space H (just as for L^2) the *Riesz-Fisher theorem* is true: if $\{\Phi_n\}$ is an arbitrary orthonormal system in H and $(c_1, c_2, \ldots, c_n, \ldots)$ is an arbitrary sequence of real numbers satisfying the condition $\sum_{k=1}^{\infty} c_k^2 < \infty$, then there is a single element X in H such that $c_k = (X, \Phi_k)$ and $\sum_{k=1}^{\infty} c_k^2 = \| X \|^2$.

The proof of this theorem differs from that of Theorem 11.6 only in that throughout the reasoning elements of H should be taken instead of those of L^2.

The Riesz-Fisher theorem allows the following fundamental theorem to be established.

Theorem 11.8. All Hilbert spaces are isomorphic to one another.

It suffices to prove that any Hilbert space H is isomorphic to l^2, and to do this it suffices to repeat the proof of Theorem 11.7, replacing throughout the elements of L^2 by those of H.

The following *statements* immediately follow from Theorem 11.8.

1°. *Any set bounded in the norm of H and containing an infinite number of elements of H is weakly compact.*

2°. *For every linear continuous functional $l(X)$ defined on the elements X of a Hilbert space H, there is one and only one element Y of H such that for all X of H $l(X) = (X, Y)$, with*

$$\| l \| = \sup_{X \in H} \frac{| l(X) |}{\| X \|} = \| Y \|.$$

Remark. It can be proved that every weakly compact set M of an infinite number of elements of H is bounded (in the norm of H). In other words, it can be proved that the *boundedness of a subset M of H containing an infinite number of elements is a necessary and sufficient condition of weak compactness of M.*

11.3.2. The equivalence of the completeness and the closure of an orthonormal system in Hilbert space. According to Theorem 10.7 in any Euclidean space (and therefore in any Hilbert space) any closed orthonormal system is complete. We now prove that the converse is also true in Hilbert space.

Theorem 11.9. Any complete orthonormal system of elements of an arbitrary Hilbert space H is closed.

Proof. Let $\{\Phi_n\}$ be an arbitrary complete orthonormal system of elements of H, and let Ψ be any element of H. It suffices to prove that the nth partial sum S_n of the Fourier series of an element Ψ with respect to $\{\Phi_n\}$ converges to that element Ψ in the norm of H.

Let $c_k = (\Psi, \Phi_k)$, $S_n = \sum_{k=1}^{n} c_k \Phi_k$. Since $\sum_{k=1}^{\infty} c_k^2$ converges* and since (by the scalar product axioms and the orthonormality of $\{\Phi_n\}$) for any $m \geqslant n$

$$\| S_m - S_n \| = \| \sum_{k=n+1}^{m} c_k \Psi_k \| = \left(\sum_{k=n+1}^{m} c_k \Psi_k, \sum_{k=n+1}^{m} c_k \Psi_k \right) = \sum_{k=n+1}^{m} c_k^2,$$

the sequence $\{S_n\}$ is *fundamental*.

But then by virtue of the completeness of H there is an element Ψ_0 of H such that

$$\| S_n - \Psi_0 \| \to 0 \text{ as } n \to \infty. \tag{11.31}$$

It remains to prove that $\Psi_0 = \Psi$. To do this it suffices to establish that the elements Ψ and Ψ_0 have the same Fourier coefficients**. Fix an arbitrary k. For any $n \geqslant k$, by the orthonormality of $\{\Phi_n\}$ and the scalar product axioms

$$(S_n, \Phi_k) = \left(\sum_{l=1}^{n} c_l \Phi_l, \Phi_k \right) = \sum_{l=1}^{n} c_l (\Phi_l, \Phi_k) = c_k. \tag{11.32}$$

On the other hand, since on the basis of the Cauchy-Buniakowski inequality

$$| (S_n, \Phi_k) - (\Psi_0, \Phi_k) | = | (S_n - \Psi_0, \Phi_k) | \leqslant$$
$$\leqslant \sqrt{\| S_n - \Psi_0 \| \cdot \| \Phi_k \|} = \sqrt{\| S_n - \Psi_0 \|},$$

it follows from (11.31) that

$$(S_n, \Phi_k) \to (\Psi_0, \Phi_k) \text{ as } n \to \infty.$$

From this relation and from (11.32) we get $(\Psi_0, \Phi_k) = c_k = (\Psi, \Phi_k)$. Thus the theorem is proved.

Corollary. *In a Hilbert space H the completeness of an orthonormal system is equivalent to its closure.*

Remark. For an incomplete Euclidean space Theorem 11.9 is in general not true.

The following example will demonstrate this fact***.

Consider a Euclidean space C^0 of all functions $f(x)$ *continuous* on a closed interval $[-\pi, \pi]$ with a scalar product defined by

$$(f, g) = \int_{-\pi}^{\pi} f(x) g(x) \, dx.$$

---

* The convergence of that series follows, for instance, from the Bessel inequality (see Theorem 10.10).

** Indeed, the coincidence of all Fourier coefficients of Ψ and Ψ_0 would mean that the element $\Psi - \Psi_0$ is orthogonal to all Φ_n and, therefore, by the completeness of the system of Φ_n, *zero*.

*** This example is due to Sh.A. Alimov.

This space is certainly not complete* (and is not therefore a Hilbert space). We construct in that space a complete orthonormal system of elements which is not closed. We do this in two steps.

1°. We first prove that in a Hilbert space $L^2[-\pi, \pi]$ there is a complete orthonormal system $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ such that the function $\varphi_0(x)$ is *discontinuous* on $[-\pi, \pi]$ and all the functions $\varphi_n(x)$, $n = 1, 2, \ldots$, are continuous there. Set

$$\Psi_0(x) = \begin{cases} \frac{1}{\sqrt{\pi}} & \text{when} \quad 0 \leqslant x \leqslant \pi, \\ 0 & \text{when} \quad -\pi \leqslant x < 0, \end{cases} \tag{11.33}$$

$$\Psi_{2n}(x) = \frac{\sqrt{2}\cos nx}{\sqrt{\pi}} \quad (n = 1, 2, \ldots),$$

$$\Psi_{2n-1}(x) = \begin{cases} \frac{\sqrt{2}\sin nx}{\sqrt{\pi}} & \text{when} \quad -\pi \leqslant x \leqslant 0, \\ 0 & \text{when} \quad 0 \leqslant x \leqslant \pi \end{cases} \qquad (n = 1, 2, \ldots).$$

Notice at once that the function $\Psi_0(x)$ is *discontinuous* on $[-\pi, \pi]$ and all the other functions $\Psi_n(x)$ $(n = 1, 2, \ldots)$ are continuous there. Besides, it is easy to verify that $\Psi_0(x)$ is orthogonal on $[-\pi, \pi]$ to each of $\Psi_n(x)$ (for all $n = 1, 2, \ldots$).

We show that although the system $\{\Psi_n(x)\}$ $(n = 0, 1, 2, \ldots)$ is not a system orthonormal in $L^2[-\pi, \pi]$ nevertheless it is comlete in the sense that any element $f(x)$ of $L^2[-\pi, \pi]$ orthogonal to every $\Psi_n(x)$ (with $n = 0, 1, 2, \ldots$) is equivalent to identical zero.

Indeed, let $f(x)$ be any element of $L^2[-\pi, \pi]$ orthogonal to every $\Psi_n(x)$ $(n = 0, 1, 2, \ldots)$.

From the orthogonality of $f(x)$ to every element of $\{\Psi_{2n-1}(x)\}$ $(n=1, 2, \ldots)$ *it follows that on* $[-\pi, 0]$ $f(x)$ *is orthogonal to* a system $\left\{\frac{\sqrt{2}\sin nx}{\sqrt{\pi}}\right\}$ $(n=1, 2, \ldots)$ and therefore, by the completeness of that system on $[-\pi, 0]$ (established in Remark 1 of Section 10.3.2), $f(x)$ is equivalent to zero on $[-\pi, 0]$.

In such a case the orthogonality of $f(x)$ to every element $\Psi_{2n}(x)$ $(n=0, 1, 2, \ldots)$ implies that on $[0, \pi]$ $f(x)$ is orthogonal to a system $\frac{1}{\sqrt{\pi}}, \frac{\sqrt{2}\cos nx}{\sqrt{\pi}}$ $(n=1, 2, \ldots)$ and, by the completeness of that system on $[0, \pi]$ (established in the same Remark 1) $f(x)$ is equivalent to zero on $[0, \pi]$ too.

* It suffices to fix some function $f_0(x)$ piecewise continuous (but not strictly continuous) on $[-\pi, \pi]$ and notice that (by Corollary 2 of Section 10.3.3) the sequence of partial sums of the Fourier trigonometric series of $f_0(x)$ converges to $f_0(x)$ in the norm of $L^2[-\pi, \pi]$. On the basis of the completeness of $L^2[-\pi, \pi]$ that sequence is fundamental. Although each element of the sequence is a function continuous on $[-\pi, \pi]$ its limit in $L^2[-\pi, \pi]$, the function $f_0(x)$, is not in C^0.

Thus $f(x)$ is equivalent to zero on the whole interval $[-\pi, \pi]$.

So the system $\{\Psi_n(x)\}$ $(n=0, 1, 2, \ldots)$ is complete in $L^2[-\pi, \pi]$. Applying the orthogonalization process to the system $\Psi_1, \Psi_2, \Psi_3, \ldots, \Psi_n, \ldots$, we obtain an orthonormal system $\overline{\Psi}_0, \overline{\Psi}_1, \overline{\Psi}_2, \ldots, \overline{\Psi}_n, \ldots$. It remains to normalize the last system, i.e. to set* $\varphi_0 = \Psi_0$, $\varphi_n = \dfrac{\overline{\Psi}_n}{\|\overline{\Psi}_n\|}$ (with $n=1, 2, \ldots$).

We obtain a complete orthonormal system $\{\varphi_n\}$ $(n = 0, 1, 2, \ldots)$ whose zero element $\varphi_0(x) = \Psi_0(x)$ is defined by formula (11.33) and is a function discontinuous on $[-\pi, \pi]$ and all the other elements, being linear combinations of continuous functions, are continuous on $[-\pi, \pi]$.

2°. We now return to the space C^0 of all functions continuous on $[-\pi, \pi]$ and prove that the system $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ is complete in C^0 but is not closed in it.

We first show that $\{\varphi_n\}$ $(n = 1, 2, \ldots)$ is complete in C^0. Let Ψ be an arbitrary element of C^0 orthogonal to every φ_n, with $n = 1, 2, \ldots$, i.e. such that

$$(\Psi, \varphi_n) = 0, \text{ with } n = 1, 2, \ldots \tag{11.34}$$

Then the function

$$f = \Psi - \varphi_0 (\Psi, \varphi_0) \tag{11.35}$$

is an element of $L^2[-\pi, \pi]$ and satisfies the condition**

$$(f, \varphi_n) = 0 \text{ for all } n = 0, 1, 2, \ldots \tag{11.36}$$

By comleteness of $\{\varphi_n\}$ $(n = 0, 1, 2, \ldots)$ in $L^2[-\pi, \pi]$ it follows from (11.36) that f is a zero element, but then from (11.35) and from the fact that $\Psi(x)$ is continuous and $\varphi_0(x)$ is discontinuous on $[-\pi, \pi]$ it follows that $(\Psi, \varphi_0) = 0$. The last equation in conjunction with (11.34) implies that Ψ is a zero element, i.e. establishes the completeness of $\{\varphi_n\}$ $(n = 1, 2, \ldots)$ in C^0.

We now prove that $\{\varphi_n\}$ $(n = 1, 2, \ldots)$ is not closed in C^0. Let P be a polynomial of the form $P = \sum_{k=1}^{n} a_k\varphi_k$ with absolutely arbitrary coefficients a_k $(k = 1, 2, \ldots, n)$. By the orthonormality of $\{\varphi_n\}$ $(n = 0, 1, \ldots)$ and by the scalar product axioms

$$\|\varphi_0 - P\| = \sqrt{(\varphi_0 - P, \varphi_0 - P)} = \sqrt{\|\varphi_0\|^2 + \|P\|^2} \geqslant 1. \tag{11.37}$$

---

* We observe that $\|\Psi_0\| = 1$.

** Indeed, with $n = 1, 2, \ldots$, (11.36) follows immediately from (11.34) and from the orthogonality of φ_0 to every φ_n $(n=1, 2, \ldots)$. The equation $(f, \varphi_0) = 0$ follows from (11.35), from the scalar product axioms and from the fact that $(\varphi_0, \varphi_0) = 1$.

Since a set of continuous functions is everywhere complete in L^2 $(-\pi, \pi]$, for φ_0 there is a continuous function $f(x)$ such that

$$\| \varphi_0 - f \| < 1/2. \tag{11.38}$$

But from (11.37) and (11.38) it follows that $\| f - P \| > 1/2$ for an absolutely arbitrary polynomial (with any coefficients), and this means that an element f of C^0 cannot be approximated in the norm of L^2 $[-\pi, \pi]$ by a linear combination of elements of $\{\varphi_n\}$ ($n = 1, 2, \ldots$), i.e. that the system $\{\varphi_n\}$ ($n = 1, 2, \ldots$) is not closed in C^0.

11.4. COMPLETELY CONTINUOUS SELF-ADJOINT OPERATORS IN HILBERT SPACE

11.4.1. The linear continuous operator. Let H be an arbitrary Hilbert space. For convenience we shall label the elements of H with small Latin letters $x, y, z, \ldots$.

If there is a rule associating with each element x of H some element y of H, then we say that an *operator* A from H into H is defined and write $y = Ax$.

Definition 1. An operator A is said to be **linear** *if for any elements x and y of H and for any real numbers α and β*

$$A(\alpha x + \beta y) = \alpha \cdot Ax + \beta \cdot Ay.$$

As for the case of the functional, we shall (whenever convenient) call the elements of H *points* of H.

Definition 2. An arbitrary operator A from H into H is said to be **continuous at a point** x_0 *of H if for any sequence* $\{x_n\}$ *of elements of H converging in the norm of H to* x_0 *the corresponding sequence* $\{Ax_n\}$ *converges in the norm of H to the element* Ax_0.

Definition 3. An operator A is said to be **continuous** *if it is continuous at each point x of H.*

Definition 4. An arbitrary operator A from H into H is said to be **bounded** *if there is a constant C such that for each element x of H* $\| Ax \| \leqslant C \| x \|$.

Definitions 1 to 4 are closely similar to the corresponding definitions for the functional in Section 11.1.2.

This allows us to give without proof the following statement: *a linear operator A from H into H is continuous if and only if it is bounded.*

The proof of the statement is absolutely identical with that for Theorem 11.1.

For a linear continuous operator A (just as for a linear continuous functional) we can introduce the concept of *norm*.

Definition 5. The **norm** *of a linear continuous operator A is the supremum of the relation* $\| Ax \| / \| x \|$ *on the set of all elements* $x \neq 0$

of H (or, equivalently, the supremum of $\| Ax \|$ on the set of all elements x of H whose norm $\| x \|$ is equal to unity).

The norm of a linear continuous operator A will be designated $\| A \|$. So by definition

$$\| A \| = \sup_{\substack{\|x\|=1 \\ x\in H}} \| Ax \|. \tag{11.39}$$

Henceforth throughout this section linear continuous operators are considered.

Here is an example of a linear continuous operator in Hilbert space. Consider a Hilbert space $L^2 \, [a \leqslant t \leqslant b]$ and suppose that we are given some function of two variables, $K(t, s)$, defined and continuous in the square $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$. We prove that an integral operator A given on the elements $x(t)$ of $L^2 \, [a \leqslant t \leqslant b]$ by

$$Ax(t) = \int_a^b K(t, s)\, x(s)\, ds \tag{11.40}$$

is linear and continuous. The linearity of the operator follows immediately from the linear property of the integral.

To prove the continuity of the operator (11.40) it suffices to establish its boundedness, for which it is sufficient to establish the finiteness of its norm (11.39). We denote by M a number

$$M = \left[\int_a^b \int_a^b K^2(t, s)\, dt\, ds \right]^{1/2} \tag{11.41}$$

and show that $\| A \| \leqslant M$. By the Cauchy-Buniakowski inequality and the definition of the norm

$$| Ax(t)^2 | \leqslant \int_a^b K^2(t, s)\, ds \int_a^b x^2(s)\, ds = \| x \|^2 \int_a^b K^2(t, s)\, ds.$$

On integrating the last inequality with respect to t between a and b and using the notation (11.41) we have

$$\| Ax \| \leqslant M \, \| x \|.$$

But this implies the boundedness of A and the validity of the inequality $\| A \| < M$ for its norm. Note that for some integral operators (11.40) $\| A \|$ is exactly equal to M.

11.4.2. The adjoint operator. We now introduce an important concept of *adjoint* operator.

Suppose that in a Hilbert space H an arbitrary linear continuous operator A from H into H is given.

Fix an arbitrary element y of H and consider a functional $f(x) = f_y(x) = (Ax, y)$ (defined on all elements x of H). It is obvious that it is linear and continuous. By the Riesz theorem on the general form of a linear functional there is a unique element $h = h_y$ of H such that $f(x) = (x, h)$ for every element x of H.

So we have associated with each element y of H one and only one element h of H such that $f_y(x) = (x, h)$, i.e. we have defined in H some operator A^* such that $h = A^*y$. It is this operator A^* that is called the *operator adjoint to an operator* A.

In other words, we arrive at the following definition.

Definition 1. An operator A^* is said to be adjoint to an operator A from H into H if for any elements x and y of H

$$(Ax, y) = (x, A^*y) \tag{11.42}$$

It follows from the above that for every linear continuous operator A there is a unique adjoint operator A^*.

It is immediate from Definition 1 that if for an operator A^* there is an adjoint operator $(A^*)^*$, then $(A^*)^* = A$.

We now show that for the case where A is linear and continuous, A^* is also linear and continuous (and for A^* therefore there is an adjoint operator and $(A^*)^* = A$, the last equation allowing the operators A and A^* to be called *mutually adjoint* operators).

Theorem 11.10. An operator A^* adjoint to a linear continuous operator A is also linear and continuous, with the norms of A^* and A connected by the relation

$$\| A^* \| = \| A \|. \tag{11.43}$$

Proof. The linearity of A^* follows immediately from relation (11.42) and from the scalar product axioms. It remains to prove the boundedness of A^* and equation (11.43).

By (11.42), the relation $\| Ay \| < \| A \| \; \| y \|$* and the Cauchy-Buniakowski inequality, for *any* elements x and y of H

$$|(A^*x, y)| = |(x, Ay)| \leqslant \|x\| \cdot \|Ay\| \leqslant \|A\| \cdot \|x\| \cdot \|y\|.$$

Inserting in this inequality A^*x for y we get for *any* x of H

$$\| A^*x \|^2 = (A^*x, A^*x) \leqslant \|A\| \cdot \|x\| \cdot \|A^*x\|$$

or $\| A^*x \| \leqslant \| A \| \cdot \| x \|$.

The last inequality implies that A^* is bounded and that its norm $\| A^* \|$ satisfies the condition

$$\| A^* \| \leqslant \| A \|. \tag{11.44}$$

---

* This relation, true for any element y of H, follows from the definition of the norm of a linear continuous operator A.

The linearity and boundedness (or equivalently the continuity) of A^* ensures the existence of an adjoint operator $(A^*)^* = A$. Repeating for that operator the above reasoning we get instead of (11.44)

$$\| A \| \leqslant \| A^* \|. \tag{11.45}$$

Equations (11.44) and (11.45) yield (11.43). Thus the theorem is proved.

Definition 2. An arbitrary operator A from H into H is said to be self-adjoint *if for A there is an adjoint operator A^* coinciding with A* (i.e. if $(Ax, y) = (x, Ay)$ for any elements x and y of H).

As an illustration, consider again an integral operator (11.40) with some function $K(t, s)$ continuous on the square $[a \leqslant t \leqslant b] \times$ $\times [a \leqslant s \leqslant b]$ (this function is generally called the *kernel* of the integral operator (11.40)).

We show that the adjoint of the operator A given by equation (11.40) is the integral operator A^* given by

$$A^*x(t) = \int_a^b K(s, t)\, x(s)\, ds \tag{11.46}$$

(by $K(s, t)$ in (11.46) one should mean the same function as in (11.40) but integrated with respect to the first independent variable).

It follows from (11.40) and (11.46) that for any elements $x(t)$ and $y(t)$ of the space $L^2[a, b]$

$$(Ax, y) = \int_a^b \left(\int_a^b K(t, s)\, x(s)\, ds \right) y(t)\, dt \tag{11.47}$$

$$(x, A^*y) = \int_a^b \left(\int_a^b K(t, s)\, y(t)\, dt \right) x(s)\, ds. \tag{11.48}$$

The right-hand sides of (11.47) and (11.48) differ only in the order of integration with respect to the variables t and s and therefore coincide*. Then, so do the left-hand sides and this precisely means that the operator A^* defined by (11.46) is adjoint to the operator A defined by (11.40).

It follows from (11.40) and (11.46) that the integral operator given by (11.40) is self-adjoint if and only if $K(t, s) = K(s, t)$ for all t and s in $[a, b]$. The kernel $K(t, s)$ satisfying the equation $K(t, s) =$ $= K(s, t)$ is called *symmetric*.

We now prove the following statement.

---

* Indeed, for continuous functions $x(t)$ and $y(t)$ the equality of the right-hand sides of (11.47) and (11.48) is obvious. But then, by Theorem 11.4 and the Cauchy-Buniakowski inequality, this is true for arbitrary elements $x(t)$ and $y(t)$ of $L^2[a, b]$ as well.

Theorem 11.11. The norm $\|A\|$ of a linear continuous self-adjoint operator A is the supremum of $|(Ax, x)|$ on the set of all elements x of H, with the norm equal to unity, i.e. the norm of A is given by

$$\|A\| \sup_{\substack{\|x\|=1\\ x\in H}} |(Ax, x)|. \tag{11.49}$$

Proof. Denote by μ the quantity on the right side of (11.49) (the existence of the supremum is obvious). To prove that $\mu = \|A\|$, it suffices to establish the two inequalities $\mu \leqslant \|A\|$ and $\mu \geqslant \|A\|$. The first of these follows immediately from the fact that on the basis of the definition of the norm of an operator and the Cauchy-Buniakowski inequality for every element x for which $\|x\| = 1$

$$(Ax, x)| \leqslant \|Ax\|\cdot\|x\| = \|Ax\| \leqslant \|A\|.$$

It remains to prove the inequality $\mu \geqslant \|A\|$. Since the operator A is linear, for each element x of H*

$$|(Ax, x)| \leqslant \mu\cdot\|x\|^2. \tag{11.50}$$

Further, it follows from the scalar product axioms and from the self-adjointness of the linear operator A (i.e. from $(Ax, y) = (x, Ay)$) that for any elements x and y of H

$$4(Ax, y) = (A(x+y), x+y) - (A(x-y), x-y).$$

From this equation and from (11.50) we get

$$4|(Ax, y)| \leqslant \mu\cdot\|x+y\|^2 + \mu\|x-y\|^2 = 2\mu(\|x\|^2 + \|y\|^2).$$

From this last inequality it follows that for arbitrary elements x and y of H for which $\|x\| = \|y\| = 1$

$$|(Ax, y)| \leqslant \mu. \tag{11.51}$$

On setting in (11.51) $y = Ax/\|Ax\|$ we get, for every element x for which $\|x\| = 1$, $(Ax, Ax)/\|Ax\| \leqslant \mu$ and therefore $\|Ax\| \leqslant \mu$. Hence $\|A\| \leqslant \mu$. Thus the theorem is proved.

11.4.3. The completely continuous operator.

Definition. An operator A from H into H is said to be completely continuous *if it maps each set, bounded (in the norm), of elements of H into a compact set.*

In other words, A is said to be completely continuous if for any sequence $\{x_n\}$ of elements of H such that $\|x_n\| \leqslant C = \text{const}$ there is a subsequence $\{x_{n_k}\}$ $(k = 1, 2, \ldots)$ such that the corresponding subsequence $\{Ax_{n_k}\}$ converges in the norm of H.

Recall that a linear operator A is continuous if and only if it is

* Since for every element $x_0 = \frac{1}{\|x\|}\cdot x$ with the norm equal to unity $|(Ax_0, x_0)| \leqslant \mu$.

bounded, i.e. if and only if it maps any set bounded (in the norm of H) again into a bounded set. Since a compact set is bounded*, any completely continuous operator is continuous. It should be added that not any continuous linear operator is completely continuous. For instance, an identity operator E of the form $Ex = x$ is continuous, but it is not completely continuous: it suffices to consider the mapping of a bounded noncompact set to see this.

We prove the following lemma.

Lemma. Let A be a linear completely continuous operator from H into H. Also let $\{x_n\}$ be an arbitrary sequence of elements of H weakly converging to an element x_0 and such that $\| x_n \| = 1$ for all n. Then the sequence $\{Ax_n\}$ converges to the elements Ax_0 in the norm of H.

Proof. Since A is linear and completely continuous**, according to Section 11.4.2, there is an adjoint operator A^* and for every element x_n and an arbitrary element y $(Ax_n, y) = (x_n, A^*y)$. From this and the weak convergence of $\{x_n\}$ to x_0 we get H $(Ax_n, y) \to$ $\to (x_0, A^*y) = (Ax_0, y)$ for any element y of H as $n \to \infty$, and this means that $\{Ax_n\}$ weakly converges to Ax_0.

We now prove that $\{Ax_n\}$ converges to Ax_0 also in the norm of H.

Suppose $\{Ax_n\}$ does not converge to Ax_0 in the norm of H. Then there is $\varepsilon > 0$ such that for some subsequence of elements $\{x_{m_k}\}$ $(k = 1, 2, \ldots)$

$$\| Ax_{m_k} - Ax_0 \| \geqslant \varepsilon. \tag{11.51'}$$

Since A is completely continuous and $\| x_n \| = 1$ we can choose a subsequence $\{x_{n_p}\}$ $(p = 1, 2, \ldots)$ of $\{x_{m_k}\}$ such that the corresponding subsequence $\{Ax_{n_p}\}$ converges in the norm of H. Since by what was proved above $\{Ax_{n_p}\}$ converges weakly to Ax_0, it also converges to Ax_0 in the norm of H. But this is contrary to inequality (11.51') true for every m_k (and clearly for every n_p).

This contradiction proves the lemma.

Remark. The above lemma is a consequence of a more general statement: *an operator A from H into H is completely continuous if and only if it maps any weakly convergent sequence $\{x_n\}$ of elements of H into a sequence $\{Ax_n\}$ converging in the norm of H.*

We shall leave this statement unproved.

We now show that an integral operator A given by (11.40) (with kernel $K(t, s)$ continuous in the square $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$) is a completely continuous operator.

Let $\{x_n(t)\}$ be an arbitrary sequence of elements of $L^2[a, b]$ bounded in the norm of $L^2[a, b]$, i.e. such that for every n

$$\| x_n(t) \| \leqslant C. \tag{11.52}$$

* See Section 11.1.3.

** And therefore continuous.

It suffices to prove that the corresponding sequence of functions $y_n(t) = Ax_n(t)$ is uniformly bounded and equicontinuous on $[a, b]$. (Then by Theorem 1.12 (Arzela) we can choose a subsequence of that sequence converging uniformly on $[a, b]$ and, the more so, in the norm of $L^2[a, b]$). From (11.52) and the Cauchy-Buniakowski inequality we get

$$|y_n(t)| = \left|\int_a^b K(t, s)\, x_n(s)\, ds\right| \leqslant \left[\int_a^b K^2(t, s)\, ds\right]^{1/2} \cdot \|x_n\|$$

which proves the uniform boundedness of $\{y_n(t)\}$ on $[a, b]$*.

Notice further that it follows from the continuity and, hence, the uniform continuity of the kernel $K(t, s)$ on the square $[a \leqslant t \leqslant b] \times \times [a \leqslant s \leqslant b]$ that given an *arbitrary* $\varepsilon > 0$ we can find $\delta > 0$ such that

$$|K(t_1, s) - K(t_2, s)| < \frac{\varepsilon}{C\sqrt{b-a}} \tag{11.53}$$

for every s in $[a, b]$ and every t_1 and t_2 in $[a, b]$ such that $|t_1 - - t_2| < \delta$.

From (11.52) and (11.53) and from the Cauchy-Buniakowski inequality we get

$$|y_n(t_2) - y_n(t_1)| \leqslant \int_a^b |K(t_2, s) - K(t_1, s)| \cdot |x_n(s)|\, ds \leqslant$$

$$\leqslant \frac{\varepsilon}{C\sqrt{b-a}} \int_a^b |x_n(s)|\, ds \leqslant \frac{\varepsilon}{C\sqrt{b-a}} \cdot \|x_n\| \cdot \sqrt{\int_a^b ds} = \varepsilon$$

for every t_1 and t_2 in $[a, b]$ such that $|t_1 - t_2| < \delta$.

The last inequality proves the equicontinuity of $\{y_n(t)\}$ on $[a, b]$ and, by virtue of the foregoing, completes the proof that the operator (11.40) is completely continuous.

11.4.4. Existence of eigenvalues of a completely continuous self-adjoint operator.

Definition. A real number λ is said to be an **eigenvalue** *of an operator A if there is a nonzero element x of H satisfying the condition $Ax = \lambda x$.*

That element x is a **proper** *element of A corresponding to the eigenvalue λ.*

If A is linear, then it follows from the condition that x is a proper element of A corresponding to an eigenvalue λ that whatever the nonzero real number α the element αx is also a proper element of A

* It suffices to notice that the kernel $K(t, s)$ is continuous on the square $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$.

corresponding to the eigenvalue λ. It is natural therefore to consider all proper elements of a linear operator A to be *normalized*, i.e. to satisfy the condition $\| x \| = 1$.

The importance of the notion of proper elements lies in the fact that the action of an operator on them reduces to multiplication by some constant λ.

Not every operator A has eigenvalues*.

We prove the following main theorem.

Theorem 11.12. Every linear self-adjoint completely continuous operator A from H into H has at least one eigenvalue λ satisfying the condition $|\lambda| = \|A\|$. Of all the eigenvalues of A it is the largest in absolute value.

Proof. Denote by M and m respectively the supremum and infimum of the scalar product (Ax, x) on the set of all elements x of H satisfying $\| x \| = 1$, i.e. set

$$M = \sup_{\substack{\|x\|=1 \\ x \in H}} (Ax, x), \quad m = \inf_{\substack{\|x\|=1 \\ x \in H}} (Ax, x). \tag{11.54}$$

For definiteness we consider the case $|M| > |m|$ (the case $|M| \leqslant |m|$ can be considered quite similarly).

Since $|M| > |m|$, $M > 0$. We prove that the number $\lambda = M$ is an eigenvalue of A.

By the definition of the supremum there is a sequence $\{x_n\}$ of elements of H such that $(Ax_n, x_n) \to \lambda$ and $\| x_n \| = 1$. Since $\{x_n\}$ is bounded (in the norm of H), by the theorem on the weak compact-

* For instance, the integral operator (11.40) has no eigenvalue when $a=0$, $b=\pi$, $K(x, s) = \sum_{n=1}^{\infty} 2^{-n} \sin(n+1)x \sin ns$. Indeed, let $\varphi(x)$ be an arbitrary element of $L^2[0, \pi]$ for which $\int_0^{\pi} K(x, s)\cdot\varphi(s)\,ds = \lambda\varphi(x)$ and let $\{b_n\}$ be Fourier coefficients in the expansion of $\varphi(x)$ with respect to a *complete* system $\left\{\frac{\sqrt{2}\sin nx}{\sqrt{\pi}}\right\}$ orthonormal on $[0, \pi]$. If $\lambda=0$, then from the generalized Parseval formula $\sum_{n=1}^{\infty} 2^{-n} b_n \sin(n+1)x = 0$, whence all $b_n = 0$ and $\varphi(x) = 0$. But if $\lambda \neq 0$, then from the equation $\int_0^{\pi} K(x, s)\,\varphi(s)\,ds = \lambda\varphi(x)$ and from the properties of the kernel $K(x, s)$ ensuring uniform convergence of the Fourier series of $\varphi(x)$ we get $\sum_{n=1}^{\infty} 2^{-n} b_n \sin(n+1)x = \lambda \sum_{n=1}^{\infty} b_n \sin nx$. Since $\lambda \neq 0$, it follows from the last equation that every $b_n = 0$ and $\varphi(x) = 0$.

ness of any infinite set bounded (in the norm of H) there is a subsequence of $\{x_n\}$ weakly converging to some element x_0 of H. We renumber that subsequence, i.e. denote it again by $\{x_n\}$. So $\{x_n\}$ converges weakly to x_0 of H. But then (by the lemma of Section 11.4.3) $\{Ax_n\}$ converges to Ax_0 in the norm of H.

Since A is self-adjoint, $(Ax_n, x_0) = (x_n, Ax_0)$, from which it follows

$$(Ax_n, x_n) - (Ax_0, x_0) = (A(x_n - x_0), (x_n + x_0)). \tag{11.55}$$

Applying the Cauchy-Buniakowski inequality (11.55) yields

$$|(Ax_n, x_n) - (Ax_0, x_0)| \leqslant \| x_n + x_0 \| \cdot \| Ax_n - Ax_0 \| \to 0$$

(for $\{Ax_n\}$ converges to Ax_0 in the norm of H and $\| x_n \| = 1$).

We have thus proved that

$$(Ax_n, x_n) \to (Ax_0, x_0). \tag{11.56}$$

From this and the relation $(Ax_n, x_n) \to \lambda$ it follows that

$$(Ax_0, x_0) = \lambda. \tag{11.57}$$

We now show that $\| x_0 \| = 1$. By the Cauchy-Buniakowski inequality $|(x_n, y)| \leqslant \| x_n \| \cdot \| y \| = \| y \|$ for any element y. Proceeding in the last inequality to the limit as $n \to \infty$ and considering the weak convergence of $\{x_n\}$ to x_0 we get $|(x_0, y)| \leqslant \| y \|$ (for any element y). From this last inequality we get $\| x_0 \| \leqslant 1$ when $y = x_0$. To prove that $\| x_0 \| = 1$, it suffices to show that the assumption that $0 < \| x_0 \| < 1$ holds leads to a contradiction.

Let $0 < \| x_0 \| < 1$. Set $y_0 = x_0/\| x_0 \|$. Then $\| y_0 \| = 1$ and by the linearity of the operator and relation (11.57)

$$(Ay_0, y_0) = \frac{1}{\| x_0 \|^2}(Ax_0, x_0) = \frac{\lambda}{\| x_0 \|^2} > \lambda$$

and this (considering that $\lambda = M$) contradicts (11.54). So $\| x_0 \| = 1$.

We now prove that x_0 is a proper element corresponding to an eigenvalue λ.

Using the definition of the norm of an element, the scalar product axioms, equation (11.57), and the definition of the norm of an operator yields

$$\| Ax_0 - \lambda x_0 \|^2 = (Ax_0 - \lambda x_0, \quad Ax_0 - \lambda x_0) =$$
$$= \| Ax_0 \|^2 - 2\lambda (Ax_0, x_0) + \lambda^2 \| x_0 \|^2 = \| A \|^2 - \lambda^2.$$

By virtue of Theorem 11.11 the right-hand (and therefore the left-hand) side of the last relation is zero. But this means that $Ax_0 = \lambda x_0$, i.e. that x_0 is a proper element of the operator A corresponding to an eigenvalue λ.

In the case $|M| \leqslant |m|$ the reasoning is similar, but λ should be set equal to m.

The only thing that remains to be proved is that if there are other eigenvalues, then the eigenvalue λ satisfying the condition $|\lambda| = \|A\|$ is the largest of them in absolute value. Let λ_1 be some other eigenvalue and let x_1 be the corresponding normed proper element. Then $Ax_1 = \lambda_1 x_1$ and therefore $(Ax_1, x_1) = \lambda_1$. But then it follows immediately from the relation*

$$|\lambda| = \sup_{\substack{\|x\|=1 \\ x \in H}} |(Ax, x)|$$

that $|\lambda| \geqslant |\lambda_1|$.

Thus the proof of the theorem is complete.

Using the theorem just proved we consider the so-called *Fredholm equation of the second kind*, i.e. the relation

$$x(t) = \mu \int_a^b K(t, s)\, x(s)\, ds \tag{11.58}$$

from which given a kernel $K(t, s)$ a function $f(t)$ different from identical zero is defined as well as those values of the parameter μ for which such a function exists. Those values of μ for which there are solutions $x(t)$, not identically zero, of the integral equation (11.58) are called *eigenvalues* of that equation. Every nonzero solution of (11.58) corresponding to a given eigenvalue is an *eigenfunction* of that equation.

The inverse of an eigenvalue of the integral equation (11.58) is generally called a *characteristic number* of the equation.

Obviously, if we introduce into consideration an integral operator A given by equation (11.40), then the eigenvalues of A are characteristic numbers of the integral equation (11.58) and the corresponding proper elements of A are the eigenfunctions of (11.58).

It has been proved in Sections 11.4.1 to 11.4.3 that if the kernel $K(t, s)$ is continuous in the square $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$ and symmetric, then the operator (11.40) is linear, self-adjoint and completely continuous.

By Theorem 11.12 the integral equation (11.58) with such a kernel $K(t, s)$ has at least one characteristic number. For equation (11.58) to have at least one *eigenvalue* it should be required that it should have at least one *nonzero* characteristic number for which the condition that $K(t, s)$ does not become an identical zero** should

---

* This relation follows from (11.54) and from the fact that $\lambda = M$ when $|M| > |m|$ and $\lambda = m$ when $|M| \leqslant |m|$.

** The condition that the continuous kernel $K(t, s)$ does not become an identical zero is a necessary and sufficient condition for an integral operator A defined by (11.40) to have *nonzero* eigenvalues. Indeed, by Theorem 11.12 $\|A\| = \|\lambda\|$, where λ is the largest absolute eigenvalue of A, so that it suffices to prove that $\|A\| = 0$ if and only if $K(t, s)$ is not identically zero. If $K(t, s) =$

be added to the requirements that $K(t, s)$ should be continuous and symmetric.

So we arrive at the following *fundamental statement: if the kernel $K(t, s)$ of the Fredholm equation of the second kind* (11.58) *is continuous in the square $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$, symmetric, and is not identically zero, then the equation has at least one eigenvalue.*

Remark. We could prove that the above statement is also true if the requirement that $K(t, s)$ should be continuous on $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$ is replaced by a weaker one, that there should be a finite integral

$$\int_a^b \int_a^b K^2(t, s)\, dt\, ds.$$

(It suffices to show that when that weaker requirement holds the integral operator (11.40) from $L^2[a, b]$ into $L^2[a, b]$ continues to be completely continuous.)

11.4.5. The basic properties of the eigenvalues and proper elements of a linear completely continuous self-adjoint operator. In conclusion we discuss the basic properties of the eigenvalues and proper elements of an arbitrary linear completely continuous self-adjoint operator from H into H.

1°. *Proper elements x_1 and x_2 corresponding to two* different *eigenvalues λ_1 and λ_2 are orthogonal.*

Indeed, on the basis of the properties of a scalar product, the equations $Ax_1 = \lambda_1 x_1$, $Ax_2 = \lambda_2 x_2$, and the self-adjointness property of the operator A we get

$$(\lambda_1 - \lambda_2)(x_1, x_2) = (\lambda_1 x_1, x_2) - (x_1, \lambda_2 x_2) =$$
$$= (Ax_1, x_2) - (x_1, Ax_2) = 0.$$

Since $\lambda_1 \neq \lambda_2$, from the equation obtained it follows that $(x_1, x_2) = 0$.

2°. There may be several proper elements of A corresponding to the same eigenvalue λ. We prove, however, that *any* nonzero *eigenvalue λ may have corresponding to it only a* finite *number of linearly independent proper elements**.

Suppose that there are an infinite number of linearly independent proper elements corresponding to some $\lambda \neq 0$. On carrying out the orthogonalization and normalization process for those elements we

---

$\equiv 0$, then clearly $\| A \| = 0$. But if, conversely, $\| A \| = 0$, then the operator A given by (11.40) maps all nonzero elements of $L^2[a, b]$ into the zero element and in particular it maps into the identical zero all the elements $\{x_n(t)\}$ of some *complete* orthonormal system in $L^2[a, b]$. But this means that $K(t, s) \equiv 0$.

* The zero eigenvalue $\lambda = 0$ may have corresponding to it an infinite number of proper elements, too. For instance, for the integral operator (11.40) with $K(t, s)$ identically zero each element of some orthonormal system, $\{x_n(t)\}$ of elements of $L^2[a, b]$ is a proper element corresponding to the eigenvalue $\lambda = 0$.

obtain an infinite orthonormal system of elements $\{x_n\}$ in H, each a proper element of A corresponding to an eigenvalue $\lambda \neq 0$. Since for any element y of H the Bessel inequality $\sum\limits_{n=1}^{\infty} (x_n, y)^2 \leqslant \| y \|^2$ is true, $\lim\limits_{n \to \infty} (x_n, y) = 0 = (0, y)$, i.e. the sequence of proper elements $\{x_n\}$ weakly converges to the zero element 0. But then it follows from the condition that the operator A is completely continuous and from the lemma of Section 11.4.3 that the corresponding sequence $\{Ax_n\}$ converges in the norm of H to the element $A\,0 = 0$. By virtue of the relation $Ax_n = \lambda x_n$ we get $| \lambda | = \| Ax_n \| \to 0$ (as $n \to \infty$), which means that $| \lambda | = 0$ and contradicts the condition $\lambda \neq 0$. This contradiction proves that every $\lambda \neq 0$ may have corresponding to it only a finite number of proper elements.

The above reasoning also shows that *all proper elements* (both corresponding to the same eigenvalue λ and to different λ) *may be considered to be mutually orthogonal* (*and to have norms equal to unity*).

3°. We now prove that *if an operator A has an infinite number of eigenvalues, then any subsequence $\{\lambda_n\}$ of eigenvalues is infinitesimal.*

Let $\{\lambda_n\}$ be any sequence of eigenvalues, and let $\{x_n\}$ be the corresponding sequence of proper elements which (by virtue of the reasoning followed in proving Property 2°) may be considered to be orthonormal. Writing for any element y of H the Bessel inequality with respect to the system $\{x_n\}$ we see that $\{x_n\}$ weakly converges to the zero element. Since the operator A is completely continuous, it follows from the lemma of Section 11.4.3 that the sequence $\{Ax_n\}$ converges to the zero element in the norm of H. But then the equation $Ax_n = \lambda_n x_n$ has as a consequence the relation

$$| \lambda_n | = \| Ax_n \| \to 0 \text{ (as } n \to \infty).$$

Property 3° allows us to state that the *eigenvalues of a linear completely continuous self-adjoint operator have no other limit points on the number axis except for the zero point*.*

This means that *all eigenvalues may be numbered in the order of their non-increasing absolute values, so that*

$$| \lambda_1 | \geqslant | \lambda_2 | \geqslant | \lambda_3 | \geqslant \ldots \geqslant | \lambda_n | \geqslant \ldots,$$

with $| \lambda_n | \to 0$ as $n \to \infty$.

In particular, all the properties we have established are true for the eigenvalues and characteristic numbers of the Fredholm equation of the second kind (11.58) with a kernel $K(t, s)$ that is continuous on the square $[a \leqslant t \leqslant b] \times [a \leqslant s \leqslant b]$ and symmetric.

* For any $\varepsilon > 0$ there may be only a finite number of eigenvalues outside the interval $(-\varepsilon, \varepsilon)$.

CHAPTER 12

FUNDAMENTALS OF THE THEORY OF CURVES AND SURFACES

In this chapter we present some facts about curves and surfaces of importance for applications.

12.1. VECTOR FUNCTIONS

12.1.1. The vector function*. We introduce the concept of vector function of m variables.

*If each point M of a set $\{M\}$ of points in an m-dimensional Euclidean space E^m is assigned according to a certain law some vector r***, *then a vector function $r = r(M)$ is said to be given on* $\{M\}$, the set $\{M\}$ being *the domain* of $r = r(M)$. If $p = m$, then, as in the case of $m = 2$ or $m = 3$ (see Section 6.2.1), a vector field defined by the vector function $r(M)$ is said to be given on $\{M\}$.

The vector $r(M)$ corresponding to a given point M of $\{M\}$ will be called *a particular* (or *special*) *value of the vector function at the point M*. The collection of all particular values of $r(M)$ is called *the set of values* of the function.

If $\{M\}$ is a set of points on a given straight line and $\{u\}$ is the set of their coordinates, then $r(M)$ may obviously be considered as a vector function of a single scalar variable u:

$$r = r(u).$$

If, however, $\{M\}$ is a set of points in an m-dimensional space and $(u_1, u_2, \ldots, u_m)$ are the coordinates of a point M, then $r(M)$ is a vector function of the scalar variables $u_1, u_2, \ldots, u_m$:

$$r = r(u_1, u_2, \ldots, u_m).$$

Remark. Let $\{r_1, r_2, \ldots, r_p\}$ be the coordinates of $r(M)$. Obviously, giving a vector function $r(M)$ is equivalent to giving p scalar functions $r_1(M), r_2(M), \ldots, r_p(M)$.

Let vectors $r(M)$ be in an Euclidean space E^p. We shall assume that the initial points of those vectors coincide with the origin of the Cartesian system chosen in E^p. In this case the point set of the end points of $r(M)$ is called *the locus* of $r(M)$. The locus of a vector

* Some facts about vector functions were given in Section 5.1.6 of [1].

** The vector r is in general in a p-dimensional Euclidean space E^p and is therefore defined by the coordinates $r_1, r_2, \ldots, r_p$.

function of a single scalar variable is in general a curve. The locus of a vector function of two variables is in general a surface.

12.1.2. The limiting value of a vector function. Continuity. In close analogy with ordinary functions the concepts of limiting value and continuity can be introduced for vector functions.

As a preliminary we introduce the concepts of a convergent sequence and of limit of a sequence of vectors.

A sequence $\{a_n\}$ *is said* to converge *to a vector* a *if given any* $\varepsilon > 0$ *we can find* N *such that for* $n \geqslant N$*

$$| a_n - a | < \varepsilon.$$

The vector a *is* the limit *of* $\{a_n\}$.

In symbols

$$\lim_{n\to\infty} a_n = a.$$

Remark. If $\{a_{1n}, a_{2n}, \ldots, a_{pn}\}$ and $\{a_1, a_2, \ldots, a_p\}$ are respectively the coordinates of a_n and a, then the convergence of $\{a_n\}$ to a implies the convergence of the number sequences $\{a_{1n}\}, \{a_{2n}\}, \ldots, \{a_{pn}\}$ to the numbers $a_1, a_2, \ldots, a_p$, respectively. Also note that the convergence of the number sequence to a_1, a_2, a_p, respectively, implies the convergence of a sequence $\{a_n\}$ of vectors with coordinates $\{a_{1n}, a_{2n}, \ldots, a_{pn}\}$ to a vector a with coordinates $\{a_1, a_2, \ldots, a_p\}$. The remark follows from the following obvious inequalities**:

$$| a_{kn} - a_k | \leqslant | a_n - a | \leqslant | a_{1n} - a_1 | + | a_{2n} - a_2 | + \ldots$$
$$\ldots + | a_{pn} - a_p |.$$

Consider a vector function $r = r(M)$ defined on a set $\{M\}$ of points of an m-dimensional Euclidean space and a point A, possibly not in $\{M\}$ but having the property that in any neighbourhood of the point there is at least one point of $\{M\}$ different from A.

Definition 1. A vector b is said to be the limiting value *of a vector function* $r(M)$ *at a point* A (or the limit of $r(M)$ as $M \to A$) *if for any sequence* $M_1, M_2, \ldots, M_n, \ldots$ *of points of* $\{M\}$ *converging to* A *whose elements* M_n *are different from* A*** ($M_n \neq A$) *the corresponding sequence* $r(M_1), r(M_2), \ldots, r(M_n), \ldots$ *of the values of* $r(M)$ *converges to* b.

In symbols

$$\lim_{M\to A} r(M) = b \quad \text{or} \quad \lim_{\substack{u_1\to a_1\\ u_2\to a_2\\ \cdots\cdots\\ u_m\to a_m}} r(u_1, u_2, \ldots, u_m) = b,$$

where $a_1, a_2, \ldots, a_m$ are the coordinates of A.

---

* *The absolute value* $| a |$ of a vector a with coordinates $\{a_1, a_2, \ldots, a_p\}$ is a number $\sqrt{a_1^2 + a_2^2 + \ldots + a_p^2}$.

** The vector $a_n - a$ has coordinates $\{a_{1n} - a_1, a_{2n} - a_2, \ldots, a_{pn} - a_p\}$.

*** This requirement is accounted for, in particular, by the fact that a function $r(M)$ may or may not be defined at A.

We shall not give any definition of the limiting value of a vector function in terms of "$\varepsilon - \delta$", neither shall we give any for the case where M tends to infinity. They can be formulated in full analogy with the corresponding definitions for scalar functions.

Let a point A be in the domain of a vector function $\boldsymbol{r} = \boldsymbol{r}(M)$ and let any neighbourhood of it contain points of the domain different from A.

Definition 2. A vector function $\boldsymbol{r} = \boldsymbol{r}(M)$ *is said to be* **continuous** *at a point A if the limiting value of the function at A exists and is equal to a particular value* $\boldsymbol{r}(A)$.

A vector function $\boldsymbol{r} = \boldsymbol{r}(A)$ is said to be continuous on a set $\{M\}$ if it is continuous at each point of that set.

12.1.3. The derivative of a vector function. In Section 5.1 of [1] we discussed the derivative of a vector function of a single scalar variable. For convenience we shall formulate that notion once again.

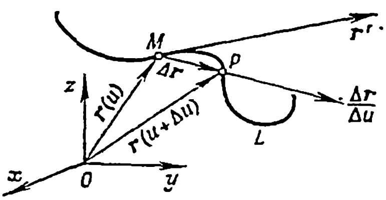


Fig. 12.1

Let $\boldsymbol{r} = \boldsymbol{r}(u)$ be a vector function of a scalar variable u. Fix a value u of the variable and assign to the variable u an arbitrary increment $\Delta u \neq 0$ such that $u + \Delta u$ is in the domain of the function. Consider a vector

$$\Delta \boldsymbol{r} = \boldsymbol{r}(u + \Delta u) - \boldsymbol{r}(u).$$

In Fig. 12.1 it coincides with the vector $\overline{MP}$. On multiplying $\Delta \boldsymbol{r}$ by a number $1/\Delta u$ we obtain a new vector

$$\frac{\Delta \boldsymbol{r}}{\Delta u} = \frac{1}{\Delta u}[\boldsymbol{r}(u + \Delta u) - \boldsymbol{r}(u)] \tag{12.1}$$

collinear with the old one. The vector (12.1) represents the average rate of change of the vector function on the interval $[u, u + \Delta u]$.

The **derivative** *of a vector function* $\boldsymbol{r} = \boldsymbol{r}(u)$ *at a given point u is the limit of the difference quotient* (12.1) *as* $\Delta u \to 0$ (provided the limit exists).

It is designated $\boldsymbol{r}'(u)$ or $\dfrac{d\boldsymbol{r}}{du}$.

Geometric considerations* make it clear that the derivative of a vector function $\boldsymbol{r} = \boldsymbol{r}(u)$ is a vector tangent to the locus of that

* They are supported by the statement in Section 12.2.2.

function. We show the relation of the derivative of $r = r(u)$ to the derivatives of its coordinates. For simplicity, consider the case where the values $r(u)$ of a vector function are vectors of a three-dimensional space. Let $\{x(u), y(u), z(u)\}$ be the coordinates of the vector function $r(u)$. Obviously, the coordinates of the difference quotient (12.1) are equal to

$$\frac{x(u+\Delta u)-x(u)}{\Delta u}, \quad \frac{y(u+\Delta u)-y(u)}{\Delta u}, \quad \frac{z(u+\Delta u)-z(u)}{\Delta u}.$$

According to the remark of Section 12.1.2 the coordinates of $r'(u)$ are equal to the derivatives $x'(u)$, $y'(u)$, $z'(u)$ of the coordinates of $r(u)$. Computing the derivative of the vector function therefore reduces to computing the derivatives of its coordinates.

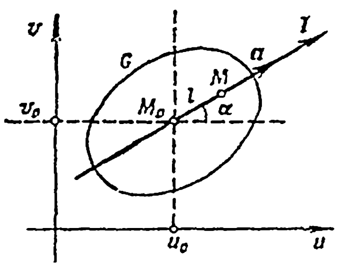


Fig. 12.2

Remark 1. The vector function $r(u)$ can be regarded as the law of motion of a particle along the locus L of the function, if the variable u is considered as time. The derivative $r'(u)$ is therefore equal to the rate of motion of the particle along the curve L.

Remark 2. Note that the rules of differentiation of the various products of vector functions (the scalar, the vector, and the triple one) are identical with those for differentiation of the products of ordinary functions. This follows from the fact that the coordinates of the derivative of a vector function are equal to the derivatives of the coordinates of the function itself and from expressing those products in terms of the coordinates of the factors.

Here are the rules of differentiation of vector function products:

$$\{r(u)\,s(u)\}' = r'(u)\,s(u) + r(u)\,s'(u),$$

$$\{[r(u)\,s(u)]\}' = [r'(u)\,s(u)] + [r(u)\,s'(u)],$$

$$\{r(u)\,s(u)\,t(u)\}' = r'(u)\,s(u)\,t(u) + r(u)\,s'(u)\,t(u) +$$
$$+ r(u)\,s(u)\,t'(u).$$

We now proceed to discuss differentiation of vector functions of several scalar variables. Since in what follows we shall use vector functions of two scalar variables u and v, we shall confine ourselves to this case.

Let a vector function $r = r(u, v)$ be given in some neighbourhood G of a point $M_0(u_0, v_0)$ (Fig. 12.2). Consider in the plane (u, v) some direction defined by the unit vector a with coordinates $\cos\alpha$, $\sin\alpha$. Draw through M_0 an axis l whose direction coincides with that of a.

take on l points $M(u, v)$ and denote by l the magnitude of a directed segment M_0M of l. The coordinates (u, v) of a point M are given by

$$u = u_0 + l\cos\alpha, \quad v = v_0 + l\sin\alpha.$$

On the axis l the function $\boldsymbol{r} = \boldsymbol{r}(u, v)$ is obviously a vector function of a single variable l. *If that function has at the point $l = 0$ a derivative with respect to the variable l, then that derivative is said to be* **the derivative** *of $\boldsymbol{r} = \boldsymbol{r}(u, v)$ at the point $M_0(u_0, v_0)$ with respect* **to the direction** *of l and designated* $\frac{\partial \boldsymbol{r}}{\partial l}$.

Remark 3. If the direction of l coincides with that of the coordinate axis u (axis v) (in Fig. 12.2 these directions are indi-

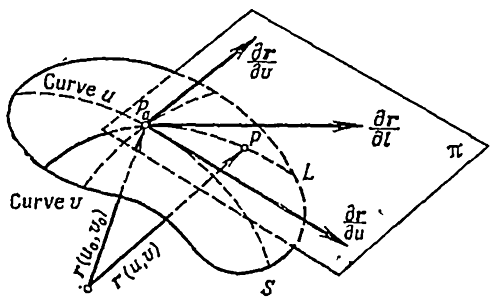


Fig. 12.3

cated by the broken lines), then the corresponding directional derivative is said to be the partial derivative of $\boldsymbol{r}(u, v)$ and designated $\frac{\partial \boldsymbol{r}}{\partial u}$ or $\boldsymbol{r}_u$ $\left(\frac{\partial \boldsymbol{r}}{\partial v} \text{ or } \boldsymbol{r}_v\right)$. If $\frac{\partial \boldsymbol{r}}{\partial u}$ is defined at every point of some neighbourhood of $M(u, v)$, then it is a vector function in that neighbourhood. It may in its turn have a partial derivative, with respect to the variable u, for instance. It is natural to call that partial derivative the second partial derivative with respect to u and designate it $\frac{\partial^2 \boldsymbol{r}}{\partial u^2}$ (or $\boldsymbol{r}_{uu}$). Other partial derivatives of various orders can be similarly defined.

The geometrical meaning of the directional derivative will be clear from the following reasoning. The locus of $\boldsymbol{r} = \boldsymbol{r}(u, v)$ is in general a surface S (Fig. 12.3). When a point $M(u, v)$ moves along l, the end point P of $\boldsymbol{r}(u, v)$ describes on S a curve L which may be regarded as the locus of a vector function of a single variable l. The derivative $\frac{\partial \boldsymbol{r}}{\partial l}$ with respect to the direction of l is therefore the vector tangent to L at the point P_0.

If the direction of l coincides with that of the coordinate axis u, then as a point M moves along the corresponding

axis passing through M_0 the end point of the vector $r(u, v)$ describes on S a curve called *a coordinate curve* u (it is designated by the broken line in Fig. 12.3). Thus $\frac{\partial r}{\partial u}$ is the vector tangent to the coordinate curve u. The partial derivative $\frac{\partial r}{\partial v}$ is the vector tangent to the coordinate curve v.

12.1.4. Differentiability of a vector function. *The increment (or total increment)* of a vector function $r = r(u, v)$ at a point $M(u, v)$ (corresponding to the increments Δu and Δv of the independent variables) is the following expression:

$$\Delta r = r(u + \Delta u, v + \Delta v) - r(u, v).$$

A vector function $r = r(u, v)$ is said to be differentiable *at a point $M(u, v)$ if its total increment at that point may be represented as*

$$\Delta r = a\,\Delta u + b\,\Delta v + \alpha\,\Delta u + \beta\,\Delta v, \tag{12.2}$$

where a and b are some vectors independent of Δu and Δv and α and β are infinitesimal vector functions as $\Delta u \to 0$ and $\Delta v \to 0$ equal to zero when $\Delta u = \Delta v = 0$**.*

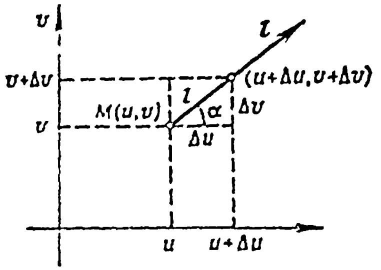


Fig. 12.4

Remark 1. If $r = r(u, v)$ is differentiable at a point $M(u, v)$, then obviously a and b are equal to $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ at the given point, respectively.

Remark 2. Let $r = r(u, v)$ be a vector function differentiable at a point $M(u, v)$ and let l be some axis passing through M in the plane (u, v) and making an angle α with the axis u. Then the derivative $\frac{\partial r}{\partial l}$ with respect to the direction of l exists and can be found from the formula

$$\frac{\partial r}{\partial l} = \frac{\partial r}{\partial u}\cos\alpha + \frac{\partial r}{\partial v}\sin\alpha. \tag{12.3}$$

Indeed, for the direction of l we have $\Delta u = l\cos\alpha$, $\Delta v = l\sin\alpha$ (Fig. 12.4). Substituting these values of Δu and Δv in relation (12.2) and using the relation $\frac{\partial r}{\partial l} = \lim\limits_{l\to 0}\frac{\Delta r}{l}$ we see that formula (12.3) is valid.

* A vector function $\alpha(\Delta u, \Delta v)$ is said to be infinitesimal if its limit as $\Delta u \to 0$ and $\Delta v \to 0$ equals zero (the zero vector).

** We give no definition of differentiability of a vector function of a single scalar variable. It can be formulated in close analogy with the corresponding definition for scalar functions of a single variable.

Remark 3. We have seen that in the case where the function $r=r(u, v)$ is differentiable formula (12.3) is true. It follows from this formula that all vectors $\frac{\partial r}{\partial l}$ are in the plane of the vectors $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$. It is natural to call the plane passing through the point of the locus of $r(u, v)$ corresponding to a point $M(u, v)$, and parallel to $\frac{\partial r}{\partial u}$ and $\frac{\partial r}{\partial v}$ *the tangential plane* to a surface S which is a locus. In Fig. 12.3 the plane π is the tangential plane to the surface S at P_0.

12.1.5. The Taylor formula for vector functions. The Taylor expansion formula for a function $r = r(u, v)$ with centre at a point $M(u, v)$ and with remainder in the Peano form is as follows:

$$r(u+\Delta u, v+\Delta v)=r(u, v)+\frac{\partial r(u, v)}{\partial u}\Delta u+\frac{\partial r(u, v)}{\partial v}\Delta v+$$
$$+\frac{1}{2!}\left(\frac{\partial^2 r(u, v)}{\partial u^2}\Delta u^2+2\frac{\partial^2 r(u, v)}{\partial u\,\partial v}\Delta u\Delta v+\right.$$
$$\left.+\frac{\partial^2 r(u, v)}{\partial v^2}\Delta v^2\right)+\ldots+\frac{1}{n!}\left(\frac{\partial^n r(u, v)}{\partial u^n}\Delta u^n+\right.$$
$$\left.+n\frac{\partial^n r(u, v)}{\partial u^{n-1}\partial v}\Delta u^{n-1}\Delta v+\ldots+\frac{\partial^n r(u, v)}{\partial v^n}\Delta v^n\right)+R_n(\Delta u, \Delta v), \quad (12.4)$$

where the remainder $R_n(\Delta u, \Delta v)$ is a vector whose order of smallness is higher than ρ^n $(\rho = \sqrt{\Delta u^2 + \Delta v^2})$*.

The validity of formula (12.4) can be shown by representing each of the coordinates of the vector $r(u, v)$ by the Taylor formula with remainder in the Peano form and then writing the expression for $r(u + \Delta u, v + \Delta v)$ with the aid of an expansion with respect to the basis vectors (the coefficients of the expansion will be the coordinates of that vector).

12.1.6. Integrals of vector functions. We have already noted that a vector function is defined by its coordinates which are scalar functions. This allows the operation of integration to be carried over to the case of vector functions.

Let a vector function $r(u)$ be given, for instance, on a closed interval $[a, b]$ and let its coordinates $r_1(u)$, $r_2(u)$, $r_3(u)$ be functions integrable on $[a, b]$. If e_1, e_2, e_3 are basis vectors, then it is natural to set by definition

$$\int_a^b r(u)\,du=e_1\int_a^b r_1(u)\,du+e_2\int_a^b r_2(u)\,du+e_3\int_a^b r_3(u)\,du.$$

* The order of smallness of a vector is defined to be the order of smallness of its absolute value.

A smooth curve L without singular points has the tangent at every point P.

We prove that the tangent is the straight line PQ passing through the point P and parallel to the vector $r'(t)$ (recall that $r'(t) \neq 0$). Indeed, the vector $\frac{\Delta r}{\Delta t}$ is parallel to the chord PM (see Fig. 12.5) and tends to $r'(t)$ as $\Delta t \to 0$. It follows that the angle between the straight line PM and the straight line PQ tends to zero as $M \to P$. Therefore the straight line PQ is the tangent to the curve L. Thus the statement is proved.

We derive a vector equation of the tangent to a curve L at a point P. Let R be the radius vector of a variable point Q on the tangent at P. The vector $\overline{PQ} = R - r(t)$ is collinear with $r'(t)$ and therefore $R - r(t) = ur'(t)$. From this we obtain the desired equation

$$R = r(t) + ur'(t) \tag{12.5}$$

where u is the parameter and t is the fixed parameter value on L defining P.

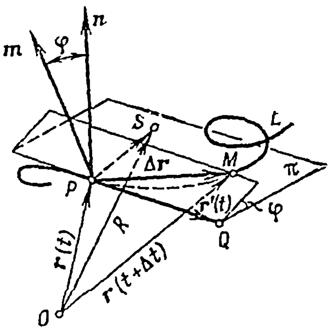


Fig. 12.6

12.2.3. The osculating plane of a curve. Let PQ be the tangent to the curve L at P (Fig. 12.6). Through PQ and the point M of L draw a plane PQM. The plane π to which PQM tends* as $M \to P$ is called *the osculating plane* to the curve L at P.

The following *statement* is true.

A regular (at least twice differentiable) curve L without singular points has the osculating plane at every point at which $r'(t)$ and $r''(t)$ are not collinear.

We prove that an osculating plane is the plane π passing through the tangent PQ and parallel to $r''(t)$. Obviously the vector

$$n = [r'(t)\, r''(t)] \tag{12.6}$$

is the normal vector to the plane π and the vector

$$m = \frac{2}{\Delta t^2}[r'(t)\, \Delta r], \quad \Delta r = r(t+\Delta t) - r(t) \tag{12.7}$$

(see Fig. 12.6) is the normal vector to the plane PQM. Since L is twice differentiable, by the Taylor formula

$$\Delta r = r'(t)\, \Delta t + \frac{1}{2} r''(t)\, \Delta t^2 + \alpha \cdot \Delta t^2, \tag{12.8}$$

* We shall say that the plane PQM tends to the plane π as $M \to P$ if the angle between them tends to zero.

where α is an infinitesimal vector function as $\Delta t \to 0$. From formulas (12.6) to (12.8) it follows that

$$m = [r'(t)\, r''(t)] + 2\,[r'(t)\, \alpha] = n + \beta, \tag{12.9}$$

where $\beta = 2\,[r'(t)\, \alpha]$ is an infinitesimal vector function as $\Delta t \to 0$. It follows from relation (12.9) that the vector m tends to n as $M \to P$ and consequently so does the angle φ between the planes PQM and π. Therefore the plane π is the osculating plane to the curve at P. Thus the statement is proved.

We derive a vector equation of an osculating plane. Let R be the radius vector of a variable point S of that plane. The vectors $\overline{PS} = R - r(t)$, $r'(t)$, and $r''(t)$ are parallel to the osculating plane and therefore $R - r(t) = ur'(t) + vr''(t)$. From this we obtain the desired equation of an osculating plane

$$R = r(t) + ur'(t) + vr''(t), \tag{12.10}$$

where u and v are the independent variables of the function R and t is the fixed parameter value on L corresponding to the point P.

We obtain the equation of an osculating plane in another form. Since $R - r(t)$, $r'(t)$, $r''(t)$ are coplanar, R satisfies the following equation:

$$(R - r(t))\, r'(t)\, r''(t) = 0. \tag{12.11}$$

If X, Y, Z are the coordinates of R (the coordinates of the variable point S of the plane π) and $x(t)$, $y(t)$, $z(t)$ are the coordinates of $r(t)$, then in coordinate form equation (12.11) becomes

$$\begin{vmatrix} X - x(t) & Y - y(t) & Z - z(t) \\ x'(t) & y'(t) & z'(t) \\ x''(t) & y''(t) & z''(t) \end{vmatrix} = 0. \tag{12.12}$$

Equation (12.12) is obviously the equation of an osculating plane.

Remark. We have defined the osculating plane geometrically by means of proceeding to the limit, and therefore if it exists it is unique. From this and from the statement proved here it follows that if at a given point π of a curve there is an osculating plane, then for any parametrization of the curve $r''(t)$ is parallel to that plane. If the parameter t is considered as time, then $r''(t)$ is the acceleration vector as the point moves along the curve L according to the law $r(t)$. Thus with any method of motion along a curve, the acceleration vector at a given point is in the osculating plane of the curve at that point. Therefore the osculating plane is also called *the plane of acceleration*.

A straight line passing through a point P of L and perpendicular to the tangent at that point is called *a normal*. A normal in the osculating plane is called *the principal normal* of the curve and a normal

perpendicular to the osculating plane is *the binormal* to the curve. Derivation of equations of these straight lines will be left to the reader.

12.2.4. The curvature of a curve. Let P be a point of a regular curve L without singular points and let M be a point of L different from P. Denote by φ the angle between the tangents at P and M and by l the length of an arc PM* (Fig. 12.7).

The curvature k_1 of L at P is the limit of the ratio φ/l as $l \to 0$ (i.e. as $M \to P$).

The following *statement* is true.

A regular (twice differentiable) curve L without singular points has a well-defined curvature k_1 at each point.

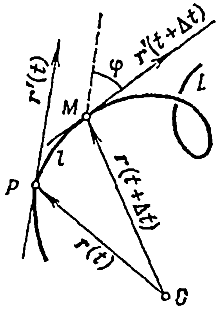


Fig. 12.7

We proceed to prove the statement. Let P and M correspond to parameter values t and $t + \Delta t$, respectively.

Evaluate $\sin\varphi$ and l. Since L is regular, $\mathbf{r}'(t) \neq 0$ at any point of L and therefore

$$\sin\varphi = \frac{|[\mathbf{r}'(t)\,\mathbf{r}'(t+\Delta t)]|}{|\mathbf{r}'(t)|\,|\mathbf{r}'(t+\Delta t)|}, \tag{12.13}$$

$$l = \int_t^{t+\Delta t} |\mathbf{r}'(\tau)|\,d\tau = |\mathbf{r}'(\tau^*)|\,\Delta t = |\mathbf{r}'(t)|\,\Delta t + \delta\Delta t \tag{12.14}$$

where $\delta \to 0$ as $\Delta t \to 0$.

Note that in the transformations of the expression for l we used the mean value formula for the integral and the continuity of the function $\mathbf{r}'(t)$.

We transform the expression (12.13) for $\sin\varphi$. By the Taylor formula

$$\mathbf{r}'(t+\Delta t) = \mathbf{r}'(t) + \mathbf{r}''(t)\,\Delta t + \boldsymbol{\alpha}\,\Delta t, \quad \boldsymbol{\alpha} \to 0 \quad \text{as} \quad \Delta t \to 0.$$

Using this formula makes the expression (12.13) for $\sin\varphi$ take the following form:

$$\sin\varphi = \frac{|[\mathbf{r}'(t)\,\mathbf{r}''(t)]| + \beta}{|\mathbf{r}'(t)|^2 + \gamma}\,\Delta t, \tag{12.15}$$

where $\beta \to 0$ and $\gamma \to 0$ as $\Delta t \to 0$.

Turning to formulas (12.14) and (12.15) and using, with $\varphi \neq 0$, the identity

$$\frac{\varphi}{l} = \frac{\varphi}{\sin\varphi}\,\frac{\sin\varphi}{l}$$

* Since L is regular, any arc PM of L is rectifiable.

$\left(\text{with } \varphi=0, \text{ the ratio } \frac{\varphi}{l} \text{ is equal to zero}\right)$ we get

$$\frac{\varphi}{l}=\frac{\varphi}{\sin\varphi}\,\frac{|[r'(t)\,r''(t)]|+\beta}{|r'(t)|^3+\mu}, \tag{12.16}$$

where β and μ tend to zero as $\Delta t\to 0$. Since $\varphi\to 0$ as $\Delta t\to 0$, $\frac{\varphi}{\sin\varphi}\to 1$ as $\Delta t\to 0$. Therefore it follows from (12.16) that as $\Delta t\to 0$, i.e. as $M\to P$, the limit $\frac{\varphi}{l}$ exists and equals $\frac{|[r'(t)\,r''(t)]|}{|r'(t)|^3}$. Thus the statement is proved.

So under the hypotheses of the statement the curvature k_1 exists and can be found from the formula

$$k_1=\frac{|[r'(t)\,r''(t)]|}{|r'(t)|^3}. \tag{12.17}$$

Remark. If the arc length l is chosen on a curve as a parameter, so that $r=r(l)$, then $|r'(l)|=1$ and $r''(l)$ is orthogonal to $r'(l)$*. In this case obviously formula (12.17) takes the following form:

$$k_1=|r''(l)|. \tag{12.18}$$

12.2.5. The torsion of a curve. Let P be a point of a regular curve L without singular points and let M be a point of L different from P. Denote by φ the angle between the osculating points at P and M and by l the length of an arc PM.

The absolute torsion $|k_2|$ of the curve L at P is the limit of the ratio φ/l as $l\to 0$ (i.e. as $M\to P$).

The following *statement* is true.

A regular (three times differentiable) curve L without singular points has a well-defined absolute torsion at every point where the curvature is different from zero.

We proceed to prove the statement.

Let P and M correspond to parameter values t and $t+\Delta t$ respectively. The normals to the osculating planes at P and M are defined by the vectors $[r'r'']_P$ and $[r'r'']_M$**. By the Taylor formula, consid-

---

* If the arc length is a parameter, then by virtue of the arbitrariness of l and Δl it follows from the formula $\Delta l=\int_l^{l+\Delta l}|r'(\tau)|\,d\tau$ that $|r'(l)|=1$ at any point of the curve. Differentiating $r'^2(l)=1$ we get $2r'(l)\,r''(l)=0$, i.e. $r''(l)$ is orthogonal to $r'(l)$.

** The expressions $[r'r'']_P$ and $[r'r'']_M$ mean that the vector product $[r'r'']$ is computed at points P and M respectively.

ering that $[r''r''] = 0$, we get

$$[r'r'']_M = [r'r'']_P + ([r'r''])_P \Delta t + \alpha \Delta t =$$
$$= [r'r'']_P + [r'r''']_P \Delta t + \alpha \Delta t, \qquad (12.19)$$

where $\alpha \to 0$ as $\Delta t \to 0$.

To compute the limit of φ/l as $l \to 0$ we shall need the value of the sine of the angle φ between the normals to the osculating planes at P and M. To this end we find the absolute value of the vector product of $[r'r'']_P$ and $[r'r'']_M$ and the product of the absolute values of these vectors. Using (12.19) yields

$$[[r'r'']_P [r'r'']_M] = [[r'r'']_P ([r'r'']_P + [r'r''']_P \Delta t + \alpha\Delta t)].$$

From this, employing the distributive property of a vector product and the well-known formula $[a\,[bc]] = b\,(ac) - c\,(ab)$ for a double vector product, we find

$$[[r'r'']_P [r'r'']_M] = r'_P (r'r''r''')_P \Delta t + \beta\Delta t,$$

where $\beta = [[r'r'']_P\,\alpha]$, and therefore $\beta \to 0$ as $\Delta t \to 0$. From the last expression for $[[r'r'']_P\,[r'r'']_M]$ we obtain the following formula:

$$|[[r'r'']_P [r'r'']_M]| = |r'_P|\,|(r'r''r''')_P|\,\Delta t + \gamma\Delta t, \qquad (12.20)$$

where $\gamma \to 0$ as $\Delta t \to 0$.

Similarly we obtain also the formula

$$|[r'r'']_P| \cdot |[r'r'']_M| = [r'r'']_P^2 + \mu\Delta t, \qquad (12.21)$$

where $\mu \to 0$ as $\Delta t \to 0$.

Using formulas (12.20) and (12.21) we obtain the required expression for $\sin\varphi$:

$$\sin\varphi = \frac{(|r'|\,(r'r''r''')| + \gamma)\,\Delta t}{[r'r'']^2 + \mu\Delta t}.$$

Note that in this expression the values for the derivatives of the vector function $r\,(t)$ are computed at P.

Turning to the expression (12.14) for l, using the just obtained formula for $\sin\varphi$ and the well-known limit $\frac{\varphi}{\sin\varphi} \to 1$ as $\varphi \to 0$ we see that the limit $\frac{\varphi}{l}$ as $l \to 0$ exists and equals $\frac{|(r'r''r''')|}{[r'r'']^2}$.

So under the hypotheses of the statement the absolute torsion $|k_2|$ exists and can be found from the formula

$$|k_2| = \frac{|(r'r''r''')|}{[r'r'']^2} \qquad (12.22)$$

We define *the torsion* k_2 of a curve using the equation

$$k_2 = + \frac{(r'r''r''')}{[r'r'']^2}. \tag{12.23}$$

We prove that k_2 *is independent of the choice of parametrization of a curve and is therefore a well-defined geometrical characteristic of a given curve**.

We proceed to another parametrization of a curve, with the aid of the parameter τ.

Denoting differentiation with respect to τ by a dot we obtain the following formulas by the indirect differentiation rule:

$$r' = \dot{r}\tau'$$

$$r'' = \ddot{r}\tau'^2 + \{\text{terms linearly expressible in terms of } \dot{r}\},$$

$$r''' = \dddot{r}\,\tau'^3 + \{\text{terms linearly expressible in terms of } \dot{r} \text{ and } \ddot{r}\}.$$

From these formulas follow the relations

$$(r'r''r''') = (\dot{r}\,\ddot{r}\,\dddot{r}\,)\,\tau'^6,\ [r'r'']^2 = [\dot{r}\,\ddot{r}]^2\tau'^6.$$

Thus

$$k_2 = \frac{(r'r''r''')}{[r'r'']^2} = \frac{(\dot{r}\,\ddot{r}\,\dddot{r}\,)}{[\dot{r}\,\ddot{r}]^2}$$

We have shown that k_2 is independent of the choice of parametrization of a curve.

12.2.6. Frenet's formulas. Natural equations of a curve. In Section 12.2.3 we have introduced the concepts of normal and binormal of a curve. These together with the tangent are the edges of a trihedral angle called *a natural trihedron*. Let the arc length be a parameter l on a curve L. Then $r'(l) = t$ is the tangent unit vector to L. Choose the unit vector n of the principal normal to be collinear with $r''(l)$** and take as the binormal unit vector

$$b = [tn]. \tag{12.24}$$

Thus t, n, b form a right-handed triple of vectors, i.e. $(tnb) > 0$. They are functions of the arc length. We find the expansions of the derivatives t', n', b' of these functions with respect to t, n, and b. Since $t = r'(l)$, we have $t' = r''(l)$. Therefore the vector t' is collinear with n:

$$t' = \alpha n.$$

* The absolute value $|k_2|$ is defined geometrically. Hence parametrization may influence only the sign of the expression $\frac{(r'r''r''')}{[r'r'']^2}$.

** According to the remark of Section 12.2.4 the vector $r''(l)$ is orthogonal to t and is in the osculating plane of the curve.

According to the remark of Section 12.2.4 $\alpha = k_1$ ($\alpha = |t'| = |r''(l)| = k_1$) and therefore

$$t' = k_1 n. \tag{12.25}$$

We now turn to the vector b. Since it is a unit vector, b' is orthogonal to b. We prove that b' is orthogonal to t as well. Differentiating the

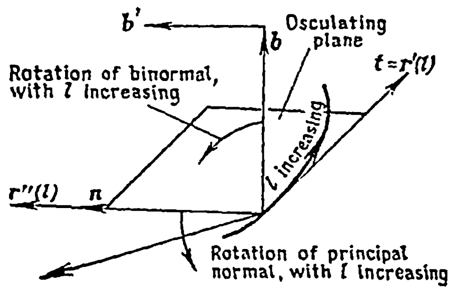


Fig. 12.8

identity $(bt) = 0$ we get $(b't) + (bt') = 0$. Since by (12.25) $(bt') = k_1(bn) = 0$, we have $(b't) = 0$ and this means that b' is orthogonal to t. It follows from the above reasoning that b' is collinear with n, i.e.

$$b' = \beta n. \tag{12.26}$$

We prove that $\beta = -k_2$. Let φ be the angle between the osculating planes of a curve at the points corresponding to parameter values l and $l + \Delta l$. Clearly the angle between $b(l)$ and $b(l+\Delta l)$ is also equal to φ, since b is orthogonal to the osculating plane. Considering therefore that $\lim\limits_{\Delta l \to 0} \frac{\varphi}{\Delta l} = k_2$ we get

$$|b'| = \lim_{\Delta l \to 0} \left| \frac{b(l+\Delta l) - b(l)}{\Delta l} \right| = \lim_{\Delta l \to 0} \left| \frac{\varphi}{\Delta l} \right| = |k_2|.$$

Consequently, since $|\beta| = |b'|$, we have the relation $|\beta| = |k_2|$. Let b' and n have the same direction. It follows from formula (12.26) that in this case $\beta = |b'|$, i.e. $\beta > 0$. It is clear that in this case $r'(l)$, $r''(l)$, and $r'''(l)$ form a triple opposite in sense to the triple t, n, b (Fig. 12.8) and therefore $(r'r''r''') < 0$, i.e. $k_2 < 0$. Since $\beta > 0$ and $|\beta| = |k_2|$, we have $\beta = -k_2$. In the case where b' and n are opposite in direction, a similar reasoning readily shows that $\beta < 0$ and $k_2 > 0$. Since $|\beta| = |k_2|$, in this case, too $\beta = -k_2$. When $\beta = 0$, the equation $\beta = -k_2$ is obvious. So we have proved that

$$\beta = -k_2. \tag{12.27}$$

From formulas (12.26) and (12.27) follows the required expression for b'

$$b' = -k_2 n. \tag{12.28}$$

We now find the expression for n'. Using the rule of vector product differentiation and formulas (12.25) and (12.28) we get

$$n' = [bt]' = [b't] + [bt'] = -k_2[nt] + k_1[bn] = -k_1 t + k_2 b.$$

Tabulating formulas (12.25), (12.28) and the expression for n' just obtained we get the following formulas called *Frenet's formulas**

$$\left.\begin{aligned} t' &= \qquad\quad k_1 n, \\ n' &= -k_1 t \qquad\quad + k_2 b, \\ b' &= \qquad -k_2 n. \end{aligned}\right\} \tag{12.29}$$

They are also called the basic formulas of the theory of curves.

It follows from Frenet's formulas that if the curvatures k_1 and k_2 of a curve L are known, it is possible to find the derivatives of the vector functions t, n, and b (i.e. the rates of change of these functions). This naturally suggests that curvature and torsion define a curve L, which is really the case. Namely, the following statement is true.

Let $k_1(l)$ and $k_2(l)$ be any differentiable functions, with $k_1(l) > 0$. Then there is a curve unique up to a position in space for which $k_1(l)$ and $k_2(l)$ are curvature and torsion respectively.

We shall not prove this statement. Note merely that its proof relies on the theorem of the existence and uniqueness of the solution of a system of ordinary differential equations.

Since according to the above statement the curvature $k_1(l)$ and torsion $k_2(l)$ completely define a curve, the system of equations

$$k_1 = k_1(l), \quad k_2 = k_2(l)$$

is usually called *the natural* (*intrinsic*) *equations* of a curve.

12.3. SOME FACTS FROM THE THEORY OF SURFACES

In Chapter 5 we have learned a number of important facts about surfaces: we were introduced to the concept of surface, to the concept of regular and smooth surface without singular points, to the concepts of tangential plane and of normal to a surface. Here we shall discuss a number of other important properties of regular surfaces.

12.3.1. The first quadratic form of a surface. Mensurations on a surface. Let Φ be a regular surface without singular points and let $r(u, v)$ be a radius vector of that surface. Then, as is known, $[r_u r_v] \neq 0$.

---

* J.F. Frenet (1816-1900) is a French mathematician

The first quadratic form I of a surface Φ is the expression

$$\mathrm{I} = d\boldsymbol{r}^2. \tag{12.30}$$

The term "quadratic form" is associated with the fact that the expression

$$I = d\boldsymbol{r}^2 = (\boldsymbol{r}_u\, du + \boldsymbol{r}_v\, dv)^2 = \boldsymbol{r}_u^2\, du^2 + 2\boldsymbol{r}_u\boldsymbol{r}_v\, du\, dv + \boldsymbol{r}_v^2\, dv^2$$

is *the quadratic form of the differentials* du and dv.

The first quadratic form is a positive definite form: it goes to 0 only if $du = dv = 0$ and is definite for the other values of du and dv. Indeed, if $d\boldsymbol{r}^2 = 0$, then $d\boldsymbol{r} = \boldsymbol{r}_u\, du + \boldsymbol{r}_v\, dv = 0$. Therefore, if du and dv do not vanish together, then from $\boldsymbol{r}_u\, du + \boldsymbol{r}_v\, dv = 0$ it follows that $\boldsymbol{r}_u$ and $\boldsymbol{r}_v$ are collinear, i.e. $[\boldsymbol{r}_u\boldsymbol{r}_v] = 0$, and this is impossible since under the hypothesis $[\boldsymbol{r}_u\boldsymbol{r}_v] \neq 0$.

For the coefficients of the first quadratic form the following notation is used:

$$\boldsymbol{r}_u^2 = E, \quad \boldsymbol{r}_u\boldsymbol{r}_v = F, \quad \boldsymbol{r}_v^2 = G. \tag{12.31}$$

In this notation the expression (12.30) for the first quadratic form becomes

$$\mathrm{I} = d\boldsymbol{r}^2 = E\, du^2 + 2F\, du\, dv + G\, dv^2. \tag{12.32}$$

So *on a regular surface Φ given by a radius vector $\boldsymbol{r} = \boldsymbol{r}(u, v)$ the first quadratic form* I *is defined by relation* (12.32). The coefficients of I can be calculated from formula (12.31).

Using the first quadratic form it is possible to make measurements on a surface: to calculate arc lengths of curves and angles between curves and to measure areas of domains.

Let L be a regular curve on Φ defined by the parametric equations*

$$u = u(t), \quad v = v(t), \quad t_0 \leqslant t \leqslant t_1, \tag{12.33}$$

$u(t)$ and $v(t)$ being differentiable functions with continuous derivatives.

---

* It is clear that representation of u and v as functions (12.33) of some parameter t defines on the surface a curve given by the vector function $\boldsymbol{r}(u(t), v(t))$. The question whether any smooth curve L on Φ can be given by parametric equations of the form (12.33) can get an affirmative answer, for instance, as follows. Let $x(t)$, $y(t)$, $z(t)$ be parametric equations of L. Then u and v as functions of t can be determined from the equations $x(t) = x(u, v)$, $y(t) = y(u, v)$, $z(t) = z(u, v)$. A solution of the form (12.33) is guaranteed by the condition $[\boldsymbol{r}_u\boldsymbol{r}_v] \neq 0$ from which it follows, for example, that $\begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \neq 0$. The last condition ensures the solvability of the system $x(t) = x(u, v)$, $y(t) = y(u, v)$ for u and v.

It is known that the length l of the arc of L defined by the radius vector $r = r(u(t), v(t))$ can be found from the formula

$$l = \int_{t_0}^{t_1} |r'(t)|\, dt \tag{12.34}$$

(see formula (11.21) in [1]).

Since $|r'(t)|\, dt = |r'(u(t), v(t))|\, dt = |dr(u, v)|$ from formula (12.34) we get

$$l = \int_{t_0}^{t_1} |r'|\, dt = \int_L |dr(u, v)| = \int_L \sqrt{dr^2} = \int_L \sqrt{\mathrm{I}} \tag{12.35}$$

(the last three integrals in (12.35) are line integrals of the first kind). So if we know the first quadratic form we can use (12.35) to calculate lengths.

We now proceed to measurements of angles on a surface.

Let a surface Φ be given by a vector function $r = r(u, v)$.

A direction $du : dv$ at a point P on Φ is defined as the direction of the vector $dr = r_u\, du + r_v\, dv$ at that point*.

Consider at P two directions, $du : dv$ and $\delta u : \delta v$. The angle φ between them is determined from the analytic-geometry formula for the cosine of the angle φ between vectors $dr = r_u\, du + r_v\, dv$ and $\delta r = r_u\, \delta u + r_v\, \delta v$:

$$\cos\varphi = \frac{(dr \cdot \delta r)}{\sqrt{dr^2}\sqrt{\delta r^2}}.$$

From this formula we obtain the following expression for $\cos\varphi$ if we take into account relations (12.31):

$$\cos\varphi = \frac{E\, du\delta u + F(du\delta v + dv\delta u) + G\, dv\, \delta v}{\sqrt{E\, du^2 + 2F\, du\, dv + G\, dv^2}\sqrt{E\delta u^2 + 2F\, \delta u\delta v + G\delta v^2}} \tag{12.36}$$

The angle between curves L_1 and L_2 on Φ intersecting in P can be defined as the angle between the directions of the tangents to L_1 and L_2 at P. Note that if a curve on a surface is defined by the parametric equations $u = u(t)$, $v = v(t)$, then the direction $du : dv$ at a point of that curve is defined by the vector

$$dr = r_u\, du + r_v\, dv = (r_u u' + r_v v')\, dt.$$

So if we know the first quadratic form we can use (12.36) to calculate the angles between directions on a surface.

The question of measuring the areas on a surface was considered in detail in Chapter 5.

Recall that if a domain Π on a surface is defined by parameters u and v in their domain Ω, then the area φ of Π can be calculated from

* Obviously, this vector is in the tangential plane at P.

the formula

$$\sigma = \iint_{\Omega} \sqrt{EG - F^2}\, du\, dv$$

(see formula (5.18)).

Thus, if we know the first quadratic form we can measure the areas of domains on a surface.

All the facts that can be obtained by measurements on a surface using the first quadratic form relate to the so-called *intrinsic geometry of surfaces.*

Two different surfaces may have the same intrinsic geometry. The simplest example of such surfaces is the plane and parabolic cylinder. Note that surfaces having the same intrinsic geometry are called *isometric.*

12.3.2. The second quadratic form of a surface. Let Φ be a regular surface defined by a radius vector $\boldsymbol{r} = \boldsymbol{r}(u, v)$ and let $\boldsymbol{n}(u, v)$ be the normal unit vector to that surface given by

$$\boldsymbol{n} = \frac{[\boldsymbol{r}_u \boldsymbol{r}_v]}{|[\boldsymbol{r}_u \boldsymbol{r}_v]|} = \frac{[\boldsymbol{r}_u \boldsymbol{r}_v]}{\sqrt{EG - F^2}}\,^{*}. \tag{12.37}$$

The second quadratic form II of a surface is the expression

$$\mathrm{II} = -d\boldsymbol{r}\, d\boldsymbol{n}. \tag{12.38}$$

Since $d\boldsymbol{r} \cdot \boldsymbol{n} = 0$**, $d(d\boldsymbol{r} \cdot \boldsymbol{n}) = 0$, i.e. $d^2\boldsymbol{r} \cdot \boldsymbol{n} = -d\boldsymbol{r}\, d\boldsymbol{n}$, and therefore the second quadratic form can also be given by

$$\mathrm{II} = d^2\boldsymbol{r} \cdot \boldsymbol{n}. \tag{12.39}$$

Since $d^2\boldsymbol{r} = \boldsymbol{r}_{uu}\, du^2 + 2\boldsymbol{r}_{uv}\, du\, dv + \boldsymbol{r}_{vv}\, dv^2$, according to (12.39) the second quadratic form may be written as

$$\mathrm{II} = (\boldsymbol{r}_{uu}\boldsymbol{n})\, du^2 + 2(\boldsymbol{r}_{uv}\boldsymbol{n})\, du\, dv + (\boldsymbol{r}_{vv}\boldsymbol{n})\, dv^2. \tag{12.40}$$

For the coefficients of the second form the following notation is used:

$$\boldsymbol{r}_{uu}\boldsymbol{n} = L, \quad \boldsymbol{r}_{uv}\boldsymbol{n} = M, \quad \boldsymbol{r}_{vv}\boldsymbol{n} = N. \tag{12.41}$$

Applying to the expression (12.37) for $\boldsymbol{n}$ and using (12.41) we obtain the following formulas for the coefficients of II:

$$L = \frac{\boldsymbol{r}_{uu}\boldsymbol{r}_u\boldsymbol{r}_v}{\sqrt{EG - F^2}}, \quad M = \frac{\boldsymbol{r}_{uv}\boldsymbol{r}_u\boldsymbol{r}_v}{\sqrt{EG - F^2}}, \quad N = \frac{\boldsymbol{r}_{vv}\boldsymbol{r}_u\boldsymbol{r}_v}{\sqrt{EG - F^2}} \tag{12.42}$$

12.3.3. Classification of points of a regular surface. We discuss deviation of a surface from the tangential plane at a given point.

---

* Since $|[\boldsymbol{r}_u \boldsymbol{r}_v]| = \sqrt{\boldsymbol{r}_u^2 \boldsymbol{r}_v^2 - (\boldsymbol{r}_u \boldsymbol{r}_v)^2}$, according to formulas (12.31) $|[\boldsymbol{r}_u \boldsymbol{r}_v]| = \sqrt{EG - F^2}$.

** The vector $d\boldsymbol{r}$ is in the tangential plane to the surface and therefore $d\boldsymbol{r} \cdot \boldsymbol{n} = 0$.

Let Φ be a regular (twice differentiable) surface, let $r = r(u, v)$ be a radius vector defining it, let $n(u, v)$ be the normal unit vector, let $P(u, v)$ be a fixed point of the surface, let n_P be the vector $n(u, v)$ at P*, and let M be a point of the surface corresponding to parameter values $u + \Delta u$, $v + \Delta v$ (Fig. 12.9).

Let N be the foot of the perpendicular dropped from M to the tangential plane π at P and let h be a quantity whose absolute value

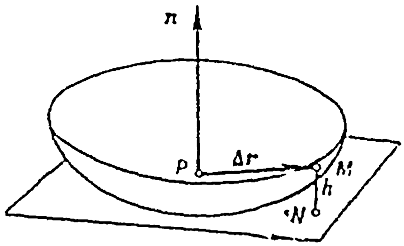


Fig. 12.9

is equal to the distance from M to π, h having a positive sign if the directions of the vectors $\overline{NM}$ and n_P coincide and a negative sign otherwise. Obviously

$$h = \Delta r \cdot n_P, \tag{12.43}$$

where $\Delta r = r(u + \Delta u, v + \Delta v) - r(u, v) = \overline{PM}$. Since u and v are independent variables, we may assume $\Delta u = du$, $\Delta v = dv$, and therefore using the Taylor formula (see formula (12.4)) we get

$$\Delta r = (dr)_P + \frac{1}{2}(d^2r)_P + R_2. \tag{12.44}$$

In this relation the differentials are calculated at P, and R_2 is a vector of an order of $o(\rho^2)$, where $\rho = \sqrt{du^2 + dv^2}$. From formulas (12.43) and (12.44) we obtain for h the following expression:

$$h = \frac{1}{2} d^2r_P \cdot n_P + R_2 n_P \tag{12.45}$$

Since $d^2r_P \cdot n_P$ is the second quadratic form II_P calculated at P and $R_2 n_P = o(\rho^2)$, relation (12.45) may be rewritten as

$$h = \frac{1}{2} \mathrm{II}_P + o(\rho^2). \tag{12.46}$$

Turning to formula (12.46) it is possible to conjecture that h is mainly influenced by the first term, $\frac{1}{2}\mathrm{II}_P$, and therefore the spatial

* In what follows a letter P at the bottom of a vector will indicate that it is taken at a point P.

structure of the surface near a regular point is defined by the second quadratic form at that point.

The following arguments support this conjecture.

1°. II_P is *of fixed sign* ($LN - M^2 > 0$).

In this case*

$$|II_P| \geqslant A\rho^2, \quad A > 0.$$

From this and from relation (12.46) it follows that h maintains sign for every sufficiently small value of ρ and therefore in a neighbourhood of a point P the surface is on one side of the tangential plane π_P at that point (Fig. 12.10).

Such a point on a surface is called *elliptical.*

The sphere, ellipsoid, elliptical paraboloid are examples of surfaces each point of which is elliptical.

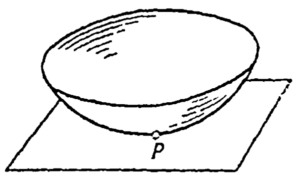


Fig. 12.10

2°. II_P is *alternating* ($LN - M^2 < 0$). In this case we can find two different directions, $du : dv$ and $\delta u : \delta v$, at a point P on a surface such that for the values of the differentials of the variables u and v corresponding to those directions II_P vanishes while all the other directions are divided by those two into two classes. For the differentials du and dv the ratio $du : dv$ of which defines the direction of one of the two classes the second form II_P is positive, and for the ratios $du : dv$ defining the directions of the other class it is negative. Therefore the surface near a point P is on either side of the tangential plane π_P at that point (Fig. 12.11).

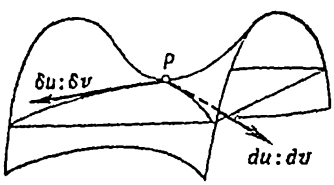


Fig. 12.11

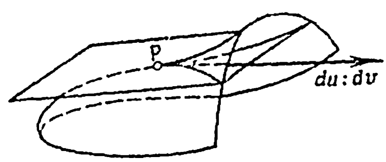


Fig. 12.12

Such a point on a surface is called *hyperbolic.*

Each point of a one-sheeted hyperboloid and of a hyperbolic paraboloid is hyperbolic.

3°. II_P is *quasi of fixed sign* ($LN - M^2 = 0$). In this case we can find a direction $du : dv$ at a point P on a surface such that for the

* It is possible, for instance, to show the validity of the inequality $|II_P| \geqslant A\rho^2$ as follows. We have $|II_P| = |L\,du^2 + 2M\,du\,dv + N\,dv^2| = |L\cos^2\alpha + 2M\cos\alpha\sin\alpha + N\sin^2\alpha|\,\rho^2$, where $\cos\alpha = du/\rho$, $\sin\alpha = dv/\rho$. Since II_P is of fixed sign, the expression $|L\cos^2\alpha + 2M\cos\alpha\sin\alpha + N\sin^2\alpha|$ has a positive minimum A, i.e. $|II_P| \geqslant A\rho^2$.

values of the differentials du and dv defining that direction the second form vanishes. For all the other values of the differentials II_p maintains sign* (Fig. 12.12).

Such a point P on a surface is called *parabolic*.

Each point of a cylinder is parabolic.

4°. II_p is equal to zero at a point P ($L = M = N = 0$). Such a point P is called *a point of flattening*. Figure 12.13 depicts a surface with a point of flattening.

Any point of a plane is a point of flattening. A point with coordinates (0, 0, 0) on a surface given by the equation $z = x^4 + y^4$ may serve as an example of an isolated point of flattening.

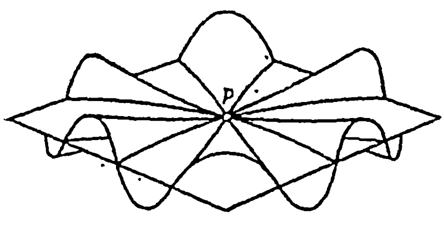


Fig. 12.13

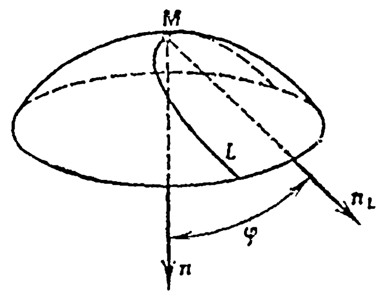


Fig. 12.14

Note that if all the points of a surface are points of flattening that surface is a plane.

12.3.4. The curvature of a curve on a surface. Let Φ be a regular surface given by a vector function $\boldsymbol{r} = \boldsymbol{r}(u, v)$, let $\boldsymbol{n}$ be the normal unit vector to Φ, and let L be a regular curve on Φ with a direction $du : dv$ at a point $P(u, v)$.

Choose the length l as the parameter on L so that $\boldsymbol{r} = \boldsymbol{r}(u(l), v(l)) = \boldsymbol{r}(l)$ along L. In Section 12.2.6 we have established that $\boldsymbol{r}''(l)$ is directed along the principal normal $\boldsymbol{n}_L$ to L at P and that the absolute value of that vector is equal to the curvature k of L at P. Therefore

$$\boldsymbol{r}''\boldsymbol{n} = k \cos \varphi, \tag{12.47}$$

where φ is the angle between the principal normal $\boldsymbol{n}_L$ to a curve L and the normal $\boldsymbol{n}$ to a surface (Fig. 12.14). By the indirect differentiation rule

$$\boldsymbol{r}''(l) = \boldsymbol{r}_{uu}u'^2 + 2\boldsymbol{r}_{uv}u'v' + \boldsymbol{r}_{vv}v'^2 + \boldsymbol{r}_u u'' + \boldsymbol{r}_v v''.$$

Since $\boldsymbol{n}$ is orthogonal to $\boldsymbol{r}_u$ and $\boldsymbol{r}_v$, substituting the obtained expression for $\boldsymbol{r}''(l)$ on the left-hand side of (12.47) and taking into account

* In this case the second form may be represented as the square of some linear form of differentials du and dv.

formulas (12.41) we get

$$r''n = (r_{uu}n)\,u'^2 + 2\,(r_{uv}n)\,u'v' + (r_{vv}n)\,v'^2 = \\ = Lu'^2 + 2Mu'v' + Nv'^2. \qquad (12.48)$$

Since $u' = \frac{du}{dl}$, $v' = \frac{dv}{dl}$ and $dl^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$ on L, from (12.47) and (12.48) we get

$$k\cos\varphi = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2} = \frac{\mathrm{II}}{\mathrm{I}}. \qquad (12.49)$$

The right-hand side of (12.49) depends only on the ratio $du : dv$, i.e. only on the direction $du : dv$. Therefore, *for all L on Φ passing through a point P in the given direction $du : dv$ the expression for $k\cos\varphi$ is equal to some constant k_n:*

$$k\cos\varphi = k_n = \text{const}. \qquad (12.50)$$

In particular, if a curve L is the so-called *normal section* L_n of a surface Φ in the direction $du : dv$, i.e. the line of intersection of Φ with the plane passing through the normal n and the direction $du : dv$, then $\varphi = 0$, $\cos\varphi = 1$ and therefore formula (12.50) becomes

$$k = k_n.$$

Thus k_n is the curvature of the normal section of a surface in the direction $du : dv$ and can be calculated from the formula

$$k_n = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2} = \frac{\mathrm{II}}{\mathrm{I}}. \qquad (12.51)$$

It is also called the normal curvature of the curve L.

Note that equation (12.50) expresses the content of *the Meusnier* theorem.*

12.3.5. Special curves on a surface.

1°. *Asymptotes.* A direction $du : dv$ on a regular surface Φ at a point P is said to be *asymptotic* if the normal curvature in that direction is equal to zero.

It follows from relation (12.51) that a direction $du : dv$ is asymptotic if and only if for that direction

$$L\,du^2 + 2M\,du\,dv + N\,dv^2 = 0. \qquad (12.52)$$

Since the second form vanishes at hyperbolic points, parabolic points and points of flattening of a surface, it is only at these points that there are asymptotic directions: two asymptotic directions at a hyperbolic point, one asymptotic direction at a parabolic point, any direction at a point of flattening being asymptotic.

We introduce the concept of *asymptote.*

---

* Meusnier (1754-1799) is a French mathematician.

An asymptote on a surface is a curve whose direction at every point is asymptotic.

If a regular surface consists of hyperbolic points, it is covered by two families of asymptotes.

For instance, the two families of rectilinear generators of a one-sheeted hyperboloid are asymptotes.

If there are two families of asymptotes on a surface, then they may in general be chosen to be coordinate curves u and v. In this case along the curve u, for instance, the parameter v remains unchanged and therefore the second form has the form $\text{II} = L\,du^2$ on u. Since in an asymptotic direction $\text{II} = 0$ (see relation (12.52)), $L = 0$. Similarly it can be shown that $N = 0$. So, *if the asymptotes of a surface are coordinate curves, then the second form is as follows:*

$$\text{II} = 2M\,du\,dv.$$

2°. *Principal directions. Lines of curvature.* It is clear from formula (12.51) that the normal curvature at a given point is a function of du and dv, more precisely, of the ratio du/dv, i.e. of the direction $du : dv$ at a given point.

Extreme values of the normal curvature at a given point are called *principal curvatures*, and the corresponding directions are *principal directions*.

We show that there are always principal directions at a given point of a regular surface.

Putting

$$\frac{du}{\sqrt{du^2+dv^2}} = \cos\alpha, \quad \frac{dv}{\sqrt{du^2+dv^2}} = \sin\alpha,$$

we transform the expression (12.51) for k_n into the form

$$k_n = \frac{L\cos^2\alpha + 2M\cos\alpha\sin\alpha + N\sin^2\alpha}{E\cos^2\alpha + 2F\cos\alpha\sin\alpha + G\sin^2\alpha}.$$

Thus the normal curvature k_n at a given point is a differentiable function of the independent variable α given on a closed interval $[0, 2\pi]$ and taking on the same values for $\alpha = 0$ and $\alpha = 2\pi$. At some interior point α of that interval therefore k_n has a local extremum. Corresponding to that value of α is a direction $du : dv$ on the surface which is naturally a principal direction. If we start measuring angles α from that principal direction, then, reasoning similarly, we shall see that an extremum of normal curvature is attained for at least one more direction $du : dv$.

So *there are at least two different principal directions at each point of a regular surface.*

We show a method of computing principal curvatures at a given point.

Assuming k_n to be a function of du and dv we obtain from (12.51) the following identity in du and dv:

$$(L-k_nE)\,du^2+2\,(M-k_nF)\,du\,dv+(N-k_nG)\,dv^2\equiv 0.$$

Differentiating it with respect to du and to dv and considering that the derivative of the normal curvature for a principal direction is equal to zero we obtain for du and dv corresponding to any principal direction the relations

$$\left.\begin{aligned}(L-k_iE)\,du+(M-k_iF)\,dv=0,\\ (M-k_iF)\,du+(N-k_iG)\,dv=0,\end{aligned}\right\} \tag{12.53}$$

where k_i is the value of a principal curvature in the direction $du: dv$. Since there are principal directions at every point, the system (12.53) has nonzero solutions for du and dv. Consequently, the determinant of (12.53) must be equal to zero:

$$\begin{vmatrix} L-k_iE & M-k_iF \\ M-k_iF & N-k_iG \end{vmatrix}=0. \tag{12.54}$$

From equation (12.54) we can determine principal curvatures k_i, and the principal directions can then be found from relations (12.53).

Equation (12.54) is a quadratic equation in k_i whose real roots are principal curvatures. Therefore two cases are possible:

1°. Equation (12.54) has two distinct roots, k_1 and k_2.

2°. The roots k_i of (12.54) coincide. We consider these cases separately.

1°. *Equation* (12.54) *has two distinct roots*: k_1 and k_2, $k_1 \neq k_2$. Two different principal directions correspond to them. We show that *if the directions of the coordinate curves u and v at a given point coincide with the principal directions, then at that point $F = 0$ and $M = 0$. Note that the vanishing of F implies that the principal directions are orthogonal.*

So let the directions of the coordinate curves u and v at a given point coincide with the principal directions. This means that the directions $du : 0$, $0 : dv$ are principal directions and therefore relations (12.53) yield

$$L-k_1E=0,\quad M-k_1F=0,$$
$$M-k_2F=0,\quad N-k_2G=0.$$

Since $k_1 \neq k_2$, clearly $M = 0$, $F = 0$. Note that with that choice of coordinate curves the principal curvatures k_1 and k_2 can be found from the relations

$$k_1=\frac{L}{E},\quad k_2=\frac{N}{G}.$$

2°. *Equation* (12.54) *has two coincident roots:* $k_1 = k_2 = k$. *We show that in this case any direction at a given point is a principal direction. If the coordinate curves at a given point are orthogonal, then at that point* $F = 0$ *and* $M = 0$.

We have already noted that there are at least *two different principal directions* at every point. In the present case to either of the principal directions corresponds the same value k of the principal curvature. But then the coefficients of the system (12.53) must vanish, i.e.

$$L - kE = 0, \quad M - kF = 0, \quad N - kG = 0.$$

From these equations it follows that at a given point the coefficients of the second form are proportional to the coefficients of the first form:

$$L = kE, \quad M = kF, \quad N = kG.$$

Substituting these values for L, M, and N in formula (12.51) we see that at a given point the curvatures of normal sections in any direction $du : dv$ are the same and equal k. Consequently, any direction $du : dv$ at a given point is a principal direction.

If the coordinate curves at a given point are orthogonal, then $F = 0$, and from the relation $M - kF = 0$ it then follows that $M = 0$, too.

So we may draw the following conclusion: *there are orthogonal principal directions at each point of a surface. If the directions of coordinate curves coincide with those principal directions, then at that point* $F = 0$ *and* $M = 0$.

We introduce the concept of *line of curvature*.

A line of curvature *on a surface is a curve whose direction at every point is principal.*

There are in general two different families of lines of curvature on any regular surface (we have pointed out above that there are two different principal directions at every point).

Note that if we choose lines of curvature as the coordinate curves, then the first and the second form of a surface will be as follows:

$$\mathrm{I} = E\,du^2 + G\,dv^2,$$

$$\mathrm{II} = L\,du^2 + N\,dv^2,$$

since $F = 0$ and $M = 0$.

3°. *Geodesics. A geodesic* (*line*) on a surface is a curve the principal normal at each point of which coincides with the normal to the surface.

Any two points of a regular complete surface can be connected by a geodesic. If the points are sufficiently close, then the geodesic connecting them is the shortest curve, any other surface curve connecting the points will be of greater length.

Note that the motion of a point on a surface without external influence takes place along a geodesic.

12.3.6. The Euler formula. The mean and the Gaussian curvature of a surface. The Gauss theorem. Let P be a fixed point of a regular surface Φ. We shall assume that coordinate curves u and v are orthogonal at the given point and that the directions of the curves coincide with the principal directions. We have established in Section 12.3.5 that under such a choice of coordinate curves at a given point

$$F = 0, \quad M = 0, \quad L - k_1 E = 0, \quad N - k_2 G = 0.$$

Using these relations makes formula (12.51) for a normal curvature k_n take the form

$$k_n = \frac{k_1 E\, du^2 + k_2 G\, dv^2}{E\, du^2 + G\, dv^2}.$$

If we set

$$\cos\varphi = \frac{\sqrt{E}\, dv}{\sqrt{E\, du^2 + G\, dv^2}}, \qquad (12.55)$$

$$\sin\varphi = \frac{\sqrt{G}\, dv}{\sqrt{E\, du^2 + G\, dv^2}},$$

then obviously we obtain the following formula for normal curvature:

$$k_n = k_1 \cos^2\varphi + k_2 \sin^2\varphi. \qquad (12.56)$$

Formula (12.56) is called *the Euler formula*. Using it a normal curvature k_n in a direction $du : dv$ can be calculated in terms of principal curvatures k_1 and k_2.

Obviously the Euler formula and formula (12.50) give complete information about the location of the curvatures of curves on a surface.

Remark 1. The angle φ in the Euler formula whose value for a given direction $du : dv$ can be found from formulas (12.55) is that angle which the direction $du : dv$ makes with the direction of the coordinate curve u.

To see this calculate from formula (12.36) the cosine of the angle beetween the direction $du : dv$ and the direction $du : 0$ of the curve u. Setting in formula (12.36) $\delta u = du$, $\delta v = 0$ yields for the desired cosine the expression $\frac{\sqrt{E}\, du}{\sqrt{E\, du^2 + G\, dv^2}}$ which coincides with the expression for $\cos\varphi$ determined from the first of the formulas (12.55).

Extensively used in the theory of surfaces are the concepts of *mean and Gaussian curvature* of a surface at a given point.

The mean curvature H *of a surface is a half-sum* $\frac{1}{2}(k_1+k_2)$ *of the principal curvatures. The* Gaussian *curvature* K *of a surface is the product* k_1k_2 *of the principal curvatures.*

Turning to equation (12.54) for principal curvatures and using the properties of the roots of a quadratic equation we obtain the following formulas for H and K:

$$H=\frac{1}{2}\,\frac{LG-2MF+NE}{EG-F^2}. \tag{12.57}$$

$$K=\frac{LN-M^2}{EG-F^2} \tag{12.58}$$

Remark 2. From the expression (12.58) for Gaussian curvature it follows that its sign coincides with that of the discriminant $LN - M^2$ of the second quadratic form (the discriminant $EG - F^2$ of the first quadratic form is always positive since the first form is positive definite). Therefore Gaussian curvature is positive at elliptical points, negative at hyperbolic points and equal to zero at parabolic points and points of flattening.

The Gaussian curvature K of a surface, on the face of it, can be found only if we know the first and second quadratic forms of the surface (see formula (12.58)).

In fact, it may be expressed in terms of the coefficients of the first quadratic form alone and is therefore the object of the intrinsic geometry of a surface. This remarkable result was established by Gauss* and is called the "*famous Gauss theorem*" in the mathematical literature. We prove this theorem.

Theorem (*Gauss*). *The Gaussian curvature K of a surface can be expressed in terms of the coefficients of the first quadratic form of the surface and their derivatives.*

Proof. Turning to formula (12.58) for Gaussian curvature K and using the expressions (12.42) for the coefficients of the second quadratic form it is easy to see that to prove the theorem it suffices to express the following expression in terms of the coefficients of the first quadratic form and their derivatives:

$$A = (r_{uu}r_ur_v)\,(r_{vv}r_ur_v) - (r_{uv}r_ur_v)^2.$$

This expression is easy to transform to the form**

$$A=\begin{vmatrix} r_{uu}r_{vv}-r_{uv}^2 & r_{uu}r_u & r_{uu}r_v \\ r_ur_{vv} & E & F \\ r_vr_{vv} & F & G \end{vmatrix} - \begin{vmatrix} 0 & r_{uv}r_u & r_{uv}r_v \\ r_ur_{uv} & E & F \\ r_vr_{uv} & F & G \end{vmatrix} \tag{12.59}$$

* Carl Friedrich Gauss (1777-1855) is a prominent German mathematician.

** The following identity is used in the transformation:

$$(a_1b_1c_1)\,(a_2b_2c_2)=\begin{vmatrix} a_1a_2 & a_1b_2 & a_1c_2 \\ b_1a_2 & b_1b_2 & b_1c_2 \\ c_1a_2 & c_1b_2 & c_1c_2 \end{vmatrix}.$$

Differentiating with respect to u and v the expression

$$\mathbf{r}_u^2=E, \quad \mathbf{r}_u\mathbf{r}_v=F, \quad \mathbf{r}_v^2=G$$

yields

$$\mathbf{r}_{uu}\mathbf{r}_u=\frac{1}{2}E_u, \quad \mathbf{r}_{uv}\mathbf{r}_u=\frac{1}{2}E_v, \; \mathbf{r}_{vv}\mathbf{r}_v=\frac{1}{2}G_v,$$

$$\mathbf{r}_{uv}\mathbf{r}_v=\frac{1}{2}G_u, \; \mathbf{r}_{uu}\mathbf{r}_v=F_u-\frac{1}{2}E_v, \; \mathbf{r}_{vv}\mathbf{r}_u=F_v-\frac{1}{2}G_u.$$

Differentiating the expression for $\mathbf{r}_{uu}\mathbf{r}_v$ with respect to v and the expression $\mathbf{r}_{uv}\mathbf{r}_v$ with respect to u and subtracting the results obtained we get

$$\mathbf{r}_{uu}\mathbf{r}_{vv}-\mathbf{r}_{uv}^2=-\frac{1}{2}G_{uu}+F_{uv}-\frac{1}{2}E_{vv}.$$

Substituting the expression obtained and the expressions for the scalar products of the derivatives on the right-hand side of (12.59) we see that the theorem is true.

In conclusion we give the expression for Gaussian curvature K in terms of the coefficients of the first quadratic form and their derivatives:

$$K=\frac{1}{(EG-F^2)^2}\begin{vmatrix}\left(-\frac{1}{2}G_{uu}+F_{uv}-\frac{1}{2}E_{vv}\right) & \frac{1}{2}E_u & \left(F_u-\frac{1}{2}E_v\right)\\ \left(F_v-\frac{1}{2}G_u\right) & E & F\\ \frac{1}{2}G_v & F & G\end{vmatrix}-$$

$$-\frac{1}{(EG-F^2)^2}\begin{vmatrix}0 & \frac{1}{2}E_v & \frac{1}{2}G_u\\ \frac{1}{2}E_v & E & F\\ \frac{1}{2}G_u & F & G\end{vmatrix}.$$

APPENDIX

EVALUATION OF FUNCTIONS FROM APPROXIMATE VALUES OF FOURIER COEFFICIENTS

A.1. The problem of summing a Fourier trigonometric series with approximate values of Fourier coefficients. Suppose first that a function $f(x)$ satisfies conditions ensuring the uniform convergence of its Fourier trigonometric series

$$\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx) \tag{A.1}$$

on the whole interval $[-\pi, \pi]$. Suppose further that instead of the exact value of the Fourier trigonometric coefficients a_k and b_k of that function we know only their approximate values $\tilde{a}_k$ and $\tilde{b}_k$. It is this case that is very common in applications.

We shall assume that errors in assigning approximate values to Fourier trigonometric coefficients are small in the sense of the norm l^2*. This means that

$$\frac{(a_0-\tilde{a}_0)^2}{2}+\sum_{k=1}^{\infty}(a_k-\tilde{a}_k)^2+(b_k-\tilde{b}_k)^2\leqslant\delta^2, \tag{A.2}$$

where δ is a sufficiently small positive number which we shall call the *error* in assigning Fourier coefficients.

The problem important for applications naturally arises: *given* approximate values for Fourier coefficients, $\tilde{a}_k$ and $\tilde{b}_k$ regenerate at a given point x, the function $f(x)$ with an error $\varepsilon(\delta)$ tending to zero as $\delta\to 0$.

We show that direct summation of a Fourier series with approximate Fourier coefficients

$$\frac{\tilde{a}_0}{2}+\sum_{k=1}^{\infty}(\tilde{a}_k\cos kx+\tilde{b}_k\sin kx) \tag{A.3}$$

cannot in general regenerate $f(x)$ at a given point x to any precision.

We fix an arbitrarily small error $\delta>0$ and set $C=\sqrt{\sum_{k=1}^{\infty}\frac{1}{k^2}}$.

Suppose that errors in assigning Fourier coefficients have the fol-

* For the definition of a space l^2 and the norm of its elements see Section 11.1.1.

lowing particular form:

$$\tilde{a}_0-a_0=0,\quad a_k-\tilde{a}_k=b_k-\tilde{b}_k=\frac{\delta}{kC\sqrt{2}},\quad \text{with } k=1, 2, \ldots$$

For Fourier coefficients with such errors relation (A.2) with the sign of exact equality will obviously be true. At the same time, replacing the exact Fourier series (A.1) by a Fourier series with approximate coefficients (A.3) causes an error equal to the sum of the series

$$\sum_{k=1}^{\infty}(\tilde{a}_k-a_k)\cos kx+(\tilde{b}_k-b_k)\sin kx$$

At the point $x=0$ the error is equal to the sum of the series

$$\sum_{k=1}^{\infty}(\tilde{a}_k-a_k)=\frac{\delta}{C\sqrt{2}}\sum_{k=1}^{\infty}\frac{1}{k}=\infty$$

(however small the error $\delta>0$ fixed).

Thus however rapidly the exact Fourier trigonometric series (A.1) may converge to $f(x)$ and however small δ in relation (A.2) giving the deviation of the approximate Fourier coefficients from the exact ones may be. direct summation of a Fourier series with approximate Fourier coefficients (A.3) cannot regenerate $f(x)$ at a given point of $[-\pi, \pi]$ to any precision.

In fact we have proved that however small the number $\delta>0$ characterizing the deviation from each other, in the sense of (A.2), of two sets of Fourier coefficients, $\{a_k, b_k\}$ and $\{\tilde{a}_k, \tilde{b}_k\}$, corresponding to the two sets. the direct sums of Fourier trigonometric series (A.1) and (A.3) may differ arbitrarily from each other.

Such problems for which an arbitrarily small deviation in initial data (in the case above these are a set of Fourier coefficients) may cause an arbitrarily large deviation of solutions corresponding to those initial data (in the case considered above, by a solution we mean the direct sum of a Fourier trigonometric series) are common in mathematics and in applications and are called *incorrectly set problems*.

In other words, the direct summation problem we have considered is an incorrectly set problem.

A general method of solving a wide class of incorrectly set problems has been developed by the Soviet mathematician A.N. Tichonoff and is called the *regularization method**.

Here we shall discuss it only as applied to the summation problem considered.

---

* A.N. Tichonoff was awarded a 1966 Lenin prize for a series of works on solving incorrectly set problems.

A.2. The regularization method for the problem of summing a Fourier trigonometric series. As applied to the problem of summing a Fourier trigonometric series with approximate Fourier coefficients the regularization method leads to an algorithm which regards as an approximate value of $f(x)$ not the sums of the series (A.3) but the sums of the series

$$\frac{\tilde{a}_0}{2}+\sum_{k=1}^{\infty}(\tilde{a}_k\cos kx+\tilde{b}_k\sin kx)\cdot\frac{1}{1+k^2\alpha} \tag{A.4}$$

obtained by multiplying the kth term of the series (A.3) by a "regularizing" factor $\frac{1}{1+k^2\alpha}$, where α is a parameter of the same order of smallness as δ in relation (A.2) giving the deviation of Fourier coefficients.

To substantiate the above algorithm we shall prove the following *main* theorem.

Theorem (A.N. Tichonoff's theorem). Let a function $f(x)$ belong to a class $L^2[-\pi, \pi]$ and be continuous at a given point x of $[-\pi, \pi]$. Then for every $\delta > 0$ and for α having the same order of smallness as δ the sum of the series (A.4) with coefficients $\tilde{a}_k$ and $\tilde{b}_k$ satisfying relation (A.2) coincides at the point x with $f(x)$ with error $\varepsilon(\delta)$ tending to zero as $\delta \to 0$.*

Proof. We may assume without loss of generality that $\alpha=\delta$ (for the case $\alpha = C(\delta)\cdot\delta$, where $0 < C_1 \leqslant C(\delta) \leqslant C_2$, can be considered quite similarly). It suffices to prove that, given any $\varepsilon > 0$, we can find $\delta_0(\varepsilon) > 0$ such that at a given point x for every positive δ satisfying the condition $\delta \leqslant \delta_0$

$$\left|\frac{\tilde{a}_0}{2}+\sum_{k=1}^{\infty}[\tilde{a}_k\cos kx+\tilde{b}_k\sin kx]\cdot\frac{1}{1+k^2\delta}-f(x)\right|\leqslant\varepsilon. \tag{A.5}$$

Fix an arbitrary $\varepsilon > 0$. We shall first show that for the fixed ε there is a number $\delta_1(\varepsilon) > 0$ such that for every positive δ satisfying the condition $\delta = \delta_1(\varepsilon)$

$$\left|\frac{\tilde{a}_0-a_0}{2}+\sum_{k=1}^{\infty}[(\tilde{a}_k-a_k)\cos kx+(\tilde{b}_k-b_k)\sin kx]\times\right.$$
$$\left.\times\frac{1}{1+k^2\delta}\right|<\frac{\varepsilon}{4}. \tag{A.6}$$

To establish (A.6) it suffices to show that the sum on the left of (A.6) tends to zero as $\delta \to +0$.

---

* This theorem is a special case of the much more general statement proved by A.N. Tichonoff.

Splitting the sum on the left of (A.6) into two sums the first of which contains terms with k satisfying the condition $k < 1/\delta$ and the second all the other terms, and applying to each of the two sums the Cauchy-Buniakowski inequality we have*

$$\left|\frac{\tilde{a}_0-a_0}{2}+\sum_{k=1}^{\infty}[(\tilde{a}_k-a_k)\cos kx+(\tilde{b}_k-b_k)\sin kx]\frac{1}{1+k^2\delta}\right|\leqslant$$

$$\leqslant\sqrt{\left\{\frac{(\tilde{a}_0-a_0)^2}{2}+\sum_{k<\frac{1}{\delta}}[(\tilde{a}_k-a_k)^2+(\tilde{b}_k-b_k)^2]\right\}O\left(\frac{1}{\delta}\right)}+$$

$$+\sqrt{\sum_{k\geqslant\frac{1}{\delta}}[(\tilde{a}_k-a_k)^2+(\tilde{b}_k-b_k)^2]\cdot\sum_{k\geqslant\frac{1}{\delta}}\frac{1}{k^4\delta^2}}. \qquad (A.7)$$

Taking into account relation (A.2) and considering that

$$\sum_{k\geqslant 1/\delta}\frac{1}{k^4}=O(\delta^3)$$

(for instance, by virtue of the integral Cauchy-Maclaurin test, see inequality (13.38) in Chapter 13 of [1]) we obtain a quantity $O(\sqrt{\delta})+O(\delta^{3/2})$ on the right of (A.7).

This inequality (A.6) may be taken for granted and to establish inequality (A.5) it suffices for us to prove that for the fixed $\varepsilon>0$ there is a number $\delta_2(\varepsilon)>0$ such that for every positive δ satisfying the condition $\delta\leqslant\delta_2(\varepsilon)$

$$\left|\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx)\frac{1}{1+k^2\delta}-f(x)\right|<\frac{3}{4}\varepsilon. \qquad (A.8)$$

Since as stated $f(x)$ is continuous at a given point x, for the fixed $\varepsilon>0$ we can fix $\eta>0$ such that for every value of y satisfying the condition $|y-x|<\eta$

$$|f(y)-f(x)|<\frac{\varepsilon}{4}. \qquad (A.9)$$

Now set $\gamma=1/\sqrt{\delta}$ and consider for the fixed x and the fixed $\eta>0$ a function $v_x(y)$ defined on a half-open interval $x-\eta<y\leqslant$

---

* We also take into account the fact that $\frac{1}{1+k^2\delta}\leqslant 1$, $\frac{1}{1+k^2\delta}<\frac{1}{k^2\delta}$.

$\leqslant x - \eta + 2\pi$ by the equation*

$$v_x(y) = \begin{cases} \frac{\gamma\pi}{2} \cdot e^{-\gamma|x-y|} & \text{when} \quad x-\eta < y < x+\eta, \\ 0 & \text{when} \quad x+\eta \leqslant y \leqslant x-\eta+2\pi \end{cases} \tag{A.10}$$

and periodically with period 2π extended to the whole infinite line $-\infty < y < +\infty$.

Compute the Fourier trigonometric coefficients A_k and B_k of $v_x(y)$.

From equation (A.10) and from the periodicity condition of $v_x(y)$ with period 2π we get**

$$A_k = \frac{1}{\pi} \int_{x-\eta}^{x+\eta} v_x(y) \cos ky\, dy = \frac{\gamma}{2} \int_{x-\eta}^{x+\eta} e^{-\gamma|x-y|} \cos ky\, dy =$$

$$= \frac{\gamma}{2} \int_{-\eta}^{\eta} e^{-\gamma|t|} \cos k(t+x)\, dt = \frac{\gamma}{2} \cos kx \int_{-\eta}^{\eta} e^{-\gamma|t|} \cos kt\, dt -$$

$$- \frac{\gamma}{2} \sin kx \int_{-\eta}^{\eta} e^{-\gamma|t|} \sin kt\, dt = \gamma \cos kx \int_{0}^{\eta} e^{-\gamma t} \cos kt\, dt.$$

Further, since

$$\int_0^{\eta} e^{-\gamma t} \cos kt\, dt = \left[\frac{e^{-\gamma t}(-\gamma \cos kt + k \sin kt)}{k^2+\gamma^2}\right]\Bigg|_{t=0}^{t=\eta} = \frac{\gamma}{k^2+\gamma^2} + e^{-\gamma\eta} \cdot \sigma_k$$

where

$$\sigma_k = \frac{-\gamma \cos k\eta + k \sin k\eta}{k^2+\gamma^2} \tag{A.11}$$

considering that $\delta = 1/\gamma^2$ yields the following expression for the Fourier coefficient A_k:

$$A_k = \frac{\cos kx}{1+k^2\delta} + e^{-\gamma\eta} \cos kx \cdot \gamma \cdot \sigma_k. \tag{A.12}$$

---

* We may assume without loss of generality that $\eta < \pi$.

** We take into account the fact that all integrals of a periodic function over intervals of length, equal to its period, coincide; we make a change of variable $y = t + x$ and take into consideration that $\int_{-\eta}^{\eta} e^{-\gamma|t|} \cos kt\, dt = 2\int_0^{\eta} e^{-\gamma t} \cos kt\, dt$, $\int_{-\eta}^{\eta} e^{-\gamma|t|} \sin kt\, dt = 0$.

Quite similarly it can be established that

$$B_k = \frac{\sin kx}{1+k^2\delta} + e^{-\gamma\eta} \sin kx \cdot \gamma \cdot \sigma_k. \tag{A.13}$$

Since as stated $f(y)$ belongs to $L^2[-\pi, \pi]$ and since $v_x(y)$ belongs to the same class for any $\delta = \frac{1}{\sqrt{\gamma}} > 0$, we have the generalized Parseval formula (see equation (11.28))

$$\frac{1}{\pi} \int_{-\pi}^{\pi} v_x(y) f(y)\, dy = \frac{A_0 a_0}{2} + \sum_{k=1}^{\infty} (A_k \cdot a_k + B_k \cdot b_k). \tag{A.14}$$

From relations (A.12), (A.13), and (A.14) it follows that to prove inequality (A.8) it suffices to establish that for every sufficiently small positive δ

$$\left| \frac{1}{\pi} \int_{-\pi}^{\pi} v_x(y) f(y)\, dy - f(x) \right| < \frac{\varepsilon}{2}, \tag{A.15}$$

$$\left| \frac{a_0}{2} e^{-\gamma\eta} \gamma\sigma_0 + e^{-\gamma\eta\gamma} \sum_{k=1}^{\infty} \sigma_k (a_k \cos kx + b_k \sin kx) \right| < \frac{\varepsilon}{4}. \tag{A.16}$$

We extend $f(y)$ periodically with period 2π to the whole infinite line.

To prove inequality (A.15) notice that by the 2π-periodicity of $v_x(y)$ and $f(x)$ and by equation (A.10)

$$\frac{1}{\pi} \int_{-\pi}^{\pi} v_x(y) f(y)\, dy = \frac{1}{\pi} \int_{x-\eta}^{x-\eta+2\pi} v_x(y) f(y)\, dy =$$

$$= \frac{\gamma}{2} \int_{x-\eta}^{x+\eta} e^{-\gamma|x-y|} f(y)\, dy = f(x) \frac{\gamma}{2} \int_{x-\eta}^{x+\eta} e^{-\gamma|x-y|}\, dy +$$

$$+ \frac{\gamma}{2} \int_{x-\eta}^{x+\eta} [f(y) - f(x)]\, e^{-\gamma|x-y|}\, dy. \tag{A.17}$$

Considering that*

$$\frac{\gamma}{2} \int_{x-\eta}^{x+\eta} e^{-\gamma|x-y|}\, dy = \frac{\gamma}{2} \int_{-\eta}^{\eta} e^{-\gamma|t|}\, dt = \gamma \int_{0}^{\eta} e^{-\gamma t}\, dt = 1 - e^{-\gamma\eta}$$

---

* In computing this integral we make a change of variable $y = t + x$.

and that for any y in $[x-\eta, x+\eta]$ inequality (A.9) is true we get with the aid of relation (A.17)

$$\left|\frac{1}{\pi}\int_{-\pi}^{\pi} v_x(y) f(y)\,dy - f(x)\right| \leqslant$$

$$\leqslant e^{-\gamma\eta}|f(x)| + \frac{\varepsilon}{4}(1-e^{-\gamma\eta}) \leqslant e^{-\gamma\eta}|f(x)| + \frac{\varepsilon}{4}.$$

Since $e^{-\gamma\eta}\,|f(x)| < \varepsilon/4$ for every fixed point x and for the fixed numbers $\varepsilon > 0$ and $\eta > 0$, with all $\delta = 1/\gamma^2$ sufficiently small, relation (A.15) is proved.

It remains to prove inequality (A.16). From (A.11) it is obvious that for σ_k, at all $k = 1, 2, \ldots$, we have the estimate

$$|\sigma_k| \leqslant 2/k. \tag{A.18}$$

For σ_0 in (A.11), with all $\delta = 1/\gamma^2$ sufficiently small, we obtain the estimate

$$|\sigma_0| \leqslant 1/\gamma \leqslant 1. \tag{A.19}$$

Applying to the sum on the left of (A.16) the Cauchy-Buniakowski inequality and using the estimates (A.18) and (A.19) we get*

$$\left[\left|\frac{a_0}{2}e^{-\gamma\eta}\gamma\sigma_0 + e^{-\gamma\eta}\gamma\sum_{k=1}^{\infty}\sigma_k(a_k\cos kx + b_k\sin kx\right| \leqslant\right.$$

$$\leqslant 2e^{-\gamma\eta}\gamma\left[\frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k^2+b_k^2)\right]^{1/2}\cdot\left[1+4\sum_{k=1}^{\infty}\frac{1}{k^2}\right]^{1/2}. \tag{A.20}$$

The two sums in square brackets on the right of (A.20) are bounded by a constant (independent of δ). The boundedness of the first sum follows immediately from the Bessel inequality and that of the second has been proved in Chapter 13 of [1].

Since for any fixed $\eta > 0$, $\lim\limits_{\delta\to 0} e^{-\gamma\eta}\cdot\gamma = \lim\limits_{\delta\to 0} e^{-\frac{\eta}{\delta^2}}\frac{1}{\delta^2} = 0$, the right-hand side of (A.20) for any fixed $\varepsilon > 0$ is less than $\varepsilon/4$, with all positive δ sufficiently small. Thus the theorem is proved.

A.3. Concluding remarks on the importance of the regularization method. A. N. Tichonoff's regularization method is of great scientific significance.

Suppose that we use some instrument to measure the frequency responses of a physical process of interest. Because of the instrument's imperfection our measurements are somewhat in error.

* We majorize by unity the absolute values of $\cos kx$ and $\sin kx$.

The problem naturally arises: if we are to clarify as much as possible our idea of the process, should we improve indefinitely the instrument's precision or should we instead develop such mathematical methods of data processing which would allow us, with the measurement accuracy available, to obtain maximum information about the process.

The regularization method shows a way to such a mathematical treatment of the observed data (i.e. of Fourier coefficients) which gives us information about the physical phenomenon under study (i.e. about the desired function $f(x)$) with an error corresponding to the error in the observed data.

NAME INDEX

SUBJECT INDEX

TO THE READER

Mir Publishers welcome your comments on the content, translation and design of this book.

We would also be pleased to receive any proposals you care to make about our publications.

Our address:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

Printed in the Union of Soviet Socialist Republics